RESEARCH ARTICLE OPEN ACCESS

WSN Using Combined Scheme of Vigor Proficient Bunching and Information Rescue

DR. A.Singaravelan [1], M. Praveen Kumar [2]
Vice Principal [1], Sva College Of Polytecnic, Namakkal
Assistant Professor [2], Department of IT, Rathinam Technical Campus, Coimbatore.
Tamil Nadu - India

ABSTRACT

The two-layer network structure has been widely adopted in wireless sensor networks (WSNs) for managing sensor nodes. In such a structure, the low layer nodes communicate with their bunch head, followed by the bunch-head nodes communicating with the base station operating in either a one hop or a multi-hop manner. The main focus of node-bunching algorithms is minimizing energy consumption due to strictly limited resources in WSNs. Also, WSNs are data intensive networks with the capability of providing users with accurate data. Unfortunately, data missing is common in WSNs. In this paper, we propose a novel joint design of sensor nodes bunching and data recovery, where the WSNs is organized in a two-layer manner with our developed bunching algorithm, and then the missing data is recovered based on this two-layer structure. Furthermore, in the proposed bunching algorithm, we take both the energy-efficiency and data forecasting accuracy into consideration and investigate the tradeoff between them. This is based on the key observation that the high energy-efficiency of the network can be achieved by reducing the distances among the nodes in a bunch, while the accuracy of the forecasting results can be improved by increasing the correlation of the data stream among the nodes in a bunch. Simulation results demonstrate that our joint design outperforms the existing algorithms in terms of energy consumption and forecasting accuracy.

Keywords :—Wireless sensor networks, vigor proficient, node bunching, data forecasting

I. INTRODUCTION

Wireless Sensor Network (WSN) becomes a popular network architecture for current and future wireless communications. Particularly in recent years, WSNs have been widely applied to various practical scenarios, such as intelligent transportation, health-care monitoring, industrial manufacture, robotics, and so on [1]. In order to provide secure and cost-efficient encryption and keying for wireless sensor networks, this paper introduce a Secure Encryption and keying based on virtual energy for wireless sensor networks (WSN) [4].Furthermore, SNs will be prevailing with the emergences of intelligent applications, e.g., Smart City [2], Wearable Computing Devices [3], Tactile Internet [4], etc. A major responsibility of the WSNs is accurately sensing and collecting useful data, for example, the measurements of air quality, humidity, biomedical and chemical information, and yielding sensed big data for further analysis [5]. At the same time, cloud-computing enabled technologies, e.g., Cloud-RAN [6] and Fog-RAN [7], provide the WSNs with the leverages of computation, communication and storage resources [8], as well as a promising method to manage, process and preserve the privacy of massive aggregated data [9]. A wireless sensor node consists of multiple modules, including battery, data process units, storage, transmitter/receiver pair, and one or several sensor devices. These sensor nodes collect the information about the surrounding environment and forward it to the base station through a one-hop or multi-hop manner. As such, WSNs serve as bridges between the physical world and human societies, resulting in a cyber-physical system [10]. However, due to limited resources, sensor nodes shall cooperate with each other to carry out complicated tasks [11–13]. For example, mobile crowd-sensing has proved to be an effective and efficient way to collect and process environmental data, as well as reconstruct the spatial field of a physical quantity (e.g., traffic condition) [15, 16]. Apart from data transmissions, WSNs need to efficiently eliminate the redundant data, query the necessary data, fuse the correlated data and recover the missing data [11]. As shown in Fig. 1, the base station sends a query message to a specific bunch head to request the readings of a node. After receiving the query message, the bunch head communicates with its members to obtain the readings, and then transmits them to the base station. Each node's readings may include several attributes, such as temperature, humidity, wind speed, and so on [12]. However, it is likely that the base station cannot successfully obtain the desired data due to the hostile environment of the communications. In this case, forecasting of missing data is needed. An intuitive method is collecting all correlated data to the base station and then forecasting the data in a centralized manner through forecasting algorithms. At the same time, intensive data transmissions and processing will cause a large amount of energy consumption at sensor nodes. As WSNs are usually battery-driven with limited power supply, battery lifetime is a vital factor for long-term operations of sensor nodes. Generally, there are two strategies to extend battery life. One is to charge the battery from other energy sources, such as energy harvesting and power transfer [12]. The other one is to develop protocols for efficiently managing energy consumptions, which has been widely studied as a hot topic in the academia society. In particular, a hybrid two-layer structure has been proposed to deal with the energy efficiency issue, where the sensor nodes are divided into multiple bunches and each bunch is managed by its bunch head. This hybrid distributed method can achieve a good tradeoff between the fully centralized and distributed approaches. Furthermore, the hybrid approach consists of two modules, including a node bunching module and a missing data recovery module. We propose a distributed data recovery scheme to address the above mentioned issue. The WSNs are assumed to be managed in a two-layer structure, i.e., the network is divided into multiple bunches and the bunch heads play a role as a bridge between the sensor nodes and the base station. To be specific, we first define both the spatial Euclidean distances between nodes [11] and the distance between the reading series generated by the nodes. Then we propose a bunching algorithm, where only the nodes having similar readings and small pairwise distances can be dispatched to the same bunch. Within each bunch, the forecasting process is conducted by each bunch head. We note that over-fitting problem imposes negative effects on the data forecasting accuracy. To address this problem, we further develop a lightweight missing data forecasting algorithm for the WSNs under the case of strictly limited resources. In the simulations, we compare the proposed scheme with the existing approaches in terms of energy consumption and forecasting accuracy. The proposed scheme is the first of its kind that integrates node bunching and missing data forecasting into a unified framework. Simulation results show that the proposed scheme performs much better in terms of forecasting accuracy compared with the existing approaches.

The major contributions are summarized as follows.

1) We identify the significance of achieving high energy efficiency and forecasting accuracy to design massive connected sensor devices networks.
2) A unified framework with the novel two-layer approach is proposed to improve the forecast accuracy and energy efficiency simultaneously. This is achieved by the sensor nodes bunching phase, followed by the missing data forecasting phase.
3) We present a bunching algorithm based on the similarities in the Euclidean distances and aggregated data. In particular, a specific approach is proposed to address the outliers in the process of bunching.
4) A lightweight missing data forecasting algorithm is developed to address the over-fitting issue, thereby significantly improving the forecasting accuracy.
5) Simulation results will demonstrate the effectiveness of the proposed algorithms to design energy efficient and missing data forecasting accurate WSNs, compared with the state-of-the-art algorithms.

The remainder of the paper is organized as follows. In Section II, we summarize the related work in two aspects, i.e., node bunching algorithms and data forecasting techniques. We then propose a novel node bunching algorithm and a data forecasting approach in Section III and IV, respectively. Extensive simulations are provided in Section V to evaluate the performance of the proposed method compared with several existing approaches. Finally, we conclude this paper and illustrate our future work in Section VI

II. RELATED WORK

In this section, we review the related works in two aspects, i.e., bunching protocols and data forecasting algorithms.

A. Nodes Bunching Approaches

Nodes bunching problem in WSNs has been intensively investigated in the literatures and many classic approaches have been proposed. In particular, linked bunch algorithm (LBA) is one of the earliest bunching algorithms, which requires no central controller and is fully distributed. In LBA, the bunch heads form a backbone network and they connect with all the sensor nodes in its bunch directly This structure is very flexible to implement a wide variety of routing strategies and can be used to avoid the problem of hidden terminals. The hierarchical control bunching algorithm proposed in treats the network as a graph. A bunch is defined as a subset of vertices whose included graph is connected. Finally, a multi-tier hierarchical bunch structure is formed and it satisfies several constrains simultaneously. In the process of bunching WSNs, the authors in argue that it was very unwise to ignore the geographical information of the sensor nodes, especially for a large WSN. They then propose a novel bunching algorithm, which used geographical radius of bunch instead of logical radius. Another classic bunching algorithm, namely, LEACH, is proposed in. LEACH forms bunches based on the received signal strength and then the bunch heads serve as a bridge between the bunch members and the base station. Various applications of LEACH illustrate that it can always produce relatively good results in terms of energy efficiency and data transmission quality.

B. Data Forecasting Algorithms

Data forecasting techniques have been widely used in WSNs to reduce data transmission and improve the energy-efficiency. The authors in propose an on-mote filtering approach relying on a local multi-step assessment of sensor data with forecasting and assessing value of information. Simulation results showed that the proposed approach reduces the number of data transmissions and the energy consumption significantly. The major disadvantage of this method is that the forecasting error accumulates with the increasing of continuous missing data in which situation yesterday becomes malfunctioning. Auto-regression based approaches are also very popular and they have been used in various scenarios. They forecast the missing data of a time series by first mining the pattern underneath the time series and then using the pattern to forecast the missing data. Similar to Yesterday, the forecasting accuracy decreases significantly with the increasing of the missing data. Different from Yesterday and auto-regression based approaches, MUSCLES makes full use of the high correlations between the co-evolving time series and construct a relation between them through linear mathematic tools. Simulation results illustrate that MUSCLES outperforms Yesterday and auto-regression based approaches in terms of forecasting accuracy. However, in MUSCLES, it is hard to define the correlations between the time series and the overfitting problem is ignored. Although the above discussed approaches in both the two fields are well developed, it is very difficult to integrate the two fields into a unified framework. In the process of designing nodes bunching algorithms, the goal is to reduce energy consumption and ignore the forecasting problem which is unwise for data-centered networks. On the other hand, most of the existing forecasting algorithms are not designed for the WSNs to recover the missing data and are not designed to be energy efficient. Besides, there are also some disadvantages for the forecasting approaches such as the over-fitting problem. As far as we know, there is no existing approaches that can solve the forecasting problem in WSNs.

III. NODE BUNCHING ALGORITHM BASED ON LOCATION AND DATA CORRELATIONS

In this section, we will present a novel node bunching algorithm by considering both the locations of all the nodes and data correlations between each pair of data streams generated by the nodes. We first assume that all the nodes in the network are located in a plain area and each node has a standard radio radius r that can be adaptively changed by the nodes in some exceptional cases. Each node in the network has the capability of severing as a bunch head and the nodes take turns to be a bunch head considering that the bunch head consumes much more energy compared with the other nodes. The bunches of all the nodes need to be reconstructed when the lasting time of a round exceeds a threshold or some bunch heads cannot serve as a bunch head anymore because of limited resources.

The operation of the bunch algorithm is divided into multiple rounds. Similar to most of the existing approaches, each round of the proposed node-bunching algorithm consists of two phases: a phase of bunch head selection and bunches formation phase. In the bunch heads selection phase, the locations and residual energy of the nodes perform more important role to reduce the energy consumption of data transmission. However, data correlations take a more important role in the process of bunch formation to increase the correlations between the nodes in a bunch. The correlations among the time series have significant impact on the forecasting accuracy. Overall, the proposed approach is operated in a half distributed manner, i.e., the bunch heads selection phase is centralized and the bunch forming phase is distributed, and this is a tradeoff between the energy consumption and missing data forecasting accuracy. The base station chose the bunch heads in a centralized way to make the distribution of bunch heads reasonable. On the contrary, the bunches formation is totally distributed and each node has its own choice to improve the forecasting accuracy. In the following, we introduce bunch heads selection and bunches' formation in Section III-A and III-B, respectively.

A. Bunch Heads Selection

Considering that most distributed approaches offer no guarantee about the number and distribution of the bunch heads, in this paper, we designed a centralized bunching algorithm to choose the bunch heads. In the initial of each round, all the nodes first transmit their IDs, location information and the residual energy to the base station. After receiving all the information about the nodes, the base station first selects the top half nodes that have more residual energy as the candidates of the bunch heads to balance the energy load among all the nodes and prolong the life time of the network. We assume that the communication energy scales exactly with squared distance and our goal is minimizing the amount of energy consumption for all the non-bunch nodes to communicate with their nearest bunch head. Note that the non-bunch head nodes may not select the nearest bunch head node as its bunch head and the bunches forming process will be discussed in Section III-B.

Finding the optimal result of the bunch heads is an NP-hard problem and it is impractical to use the brute force algorithms considering that the numbers of the nodes in WSNs are very large. Authors in propose a heuristic method to solve the problem based on genetic algorithms and this algorithm. In this paper, the method in is adopted by the base station to choose the m bunch heads.

Algorithm 1: Pseudocode of bunch heads election

1: All the nodes in the network transmit their node IDs, locations and residual energy to the base station

2: Sort the residual energy of the nodes and select the top-half nodes with more energy as the candidates of bunch heads

3: Adopt the method in to choose the m final bunch heads

4: Broadcast the IDs and locations of the selected bunch heads in the network

Given the final selected bunch heads, the base station broadcasts the IDs of the nodes in the whole network. Then each node compares its own ID with the received IDs to realize whether it is a bunch head or not. Considering that transmitting all the readings of the nodes to the base station is impractical, we ignore the data correlations between the readings of the nodes when selecting the bunch heads and only the information of location and residual energy of the nodes are considered by the station. Intuitively, a straightforward solution of node bunching is to transmit all the necessary information to the base station and both bunch heads selection and bunches formation are operating in a centralized way. Then the base station broadcasts the bunching result in the network, which is very similar to LEACH. However, this pattern is impractical for our approach, since the data correlation is also taken into consideration and the readings of the nodes are of large amount, which cannot be transmitted to the base station totally. Therefore, we design a distributed approach to form the bunches, as presented in the following.

B. Bunches Forming

We first introduce the definition of trend closeness among the readings of nodes, which is a common measurement of the correlations between the time series.

Definition 1 (Spatial Closeness between Sensor Nodes):

Under the assumption that all the nodes are located in a plain regime, the spatial distance of two nodes dist is defined as the Euclidean Distance. Sensor node np is spatial close to nq if np is at worst d far from nq. Parameter d is set by the users of the network and naturally not too much larger than the radio radius r, otherwise, the nodes in a bunch cannot communicate with each other very well. The spatial closeness between sensor nodes has important affection on the selection of bunch heads.

Definition 2 (Trend Closeness between Readings of Nodes):

Each node in the network generates readings about the surrounding environment which can be treated as a time series and in most cases the readings are strongly correlated between neighboring nodes. For each node, we treat its readings as a time series and the time series is infinite. In this paper, we define the trend closeness between two time series as the Dynamic Time Wrapping (DTW) which is more robust than the Euclidean Distance. In this paper, for convenience, we compute the trend closeness based on the latest ten readings of the nodes rather than all the historical readings of the nodes.

Based on the above two definitions, we propose a novel bunches formation algorithm for WSNs considering both the spatial locations and data correlations. The pseudocode of bunches formation is presented in Algorithm 2 and will be discussed as follows. After receiving the IDs and locations of the bunch heads, each non-bunch heads node computes the distances between itself and the bunch heads and then add the bunch heads in its communication range to the candidate set Snj . The node needs to choose a bunch head in Snj as its bunch head and join

Algorithm 2: Pseudocode of bunches forming

1: The base station broadcasts the IDs and locations of the bunch heads in the network

2: Each non-bunch head node nj computes the distances between itself and the bunch heads

3: Add the bunch heads that spatial close to nj to the candidate set Snj = fch1; ch2; ; chog

4: nj communicates with the bunch heads in Snj and get the most recent s readings from each node in Snj

5: nj compute the trend closeness between its own readings and that of each node in Snj

6: Select the node cho′ with the least distance as the bunch head and send a message to cho′ to join its bunch

7: The bunch heads maintain all the information of its bunch members the corresponding bunch based on the similarities between the readings of itself and that of the bunch heads.

IV. MISSING DATA FORECASTING

In this section, we will develop the missing data forecasting algorithms for the given bunches, calculated by the algorithms presented in the previous section. Intuitively, the base station can collect all the time series from the network and then recover the missing values based on conventional data forecasting approaches, such as MUSCLES [8]. However, this method is impractical considering that energy is strictly limited and the jamming problem in the network. Therefore, we will design a distributed framework to forecast the missing values based on the node bunching approach proposed previously. In Section IV-A,

The Over-fitting Problem in Time Series Forecasting In this section, we conduct a detailed experiment to present the over-fitting problem in time series forecasting field. This is the first clear illustration that too many uncorrelated time series will decrease the forecasting accuracy significantly. Therefore, we need to select several similar time series to forecast the missing data rather than employ all the time series. In the following, we present the experiment including collecting datasets, forecasting the missing data and analyzing of the simulation results. We first collect 13 important stock price indexes from March 2nd, 2015 to February 29th, 2016 all over the world and these indexes are LNSZZS (CN), SZSE (CN), GEI (CN), HIS (HK), Nikkei (JPN), Kospi (KR), FTSE (UK), CAC40 (FR), DAX (GER), MIB (IT), SSMI (CH), DJIA (US), and Nasdaq (US). The normalized indexes are presented in Fig. 2. For convenience, we employ letters a m to represent the time series. Intuitively, we can find that each time series may be very similar with some time series and also can be very different with some other time series. Consider a as an example, a is very similar with b and c. This is reasonable, because a, b and c are all belong to China and they influence each other significantly.

A Novel Distributed Data Forecasting Algorithm

Assume that each sensor node in a cluster update its reading periodically and then send the updated reading to the cluster head when it can communicate with the cluster head. The cluster head maintain the latest 10 reading for each node to measure the similarities between the time series generated by the nodes in the cluster. In this paper, to reduce energy consumption, the cluster heads select the corrected time series through a very simple method, i.e., selecting the most similar 5 time series to forecast the missing data as discussed in Section IV-A. For some special cases, the number of the members in a cluster may less than 5 and, in this case, we use all the time series in the cluster to forecast the missing data.

V. RESULTS AND DISCUSSION

In this section, we present the simulation results and discuss the advantages and disadvantages of the proposed approach. Obviously, for the approaches of Yesterday, Auto- Regression and MUSCLES, parameter s has no affection on the performance, because all the three approaches employ LEACH-C as the node clustering method which do not contain the parameter s. As shown in Fig. 6, with the increasing of s, the MAE decreases for the propose approach and new approach always outperforms Yesterday and Auto-Regression in terms of forecasting accuracy. However, when $s < 4$, MUSCLES performs better than the new approach and, when $4 < s < 10$, the new approach performs better than MUSCLES. In addition, when $s > 7$, the MAE of the new approach performs stable and we set $s = 8$ in the following to decrease the energy consumption.

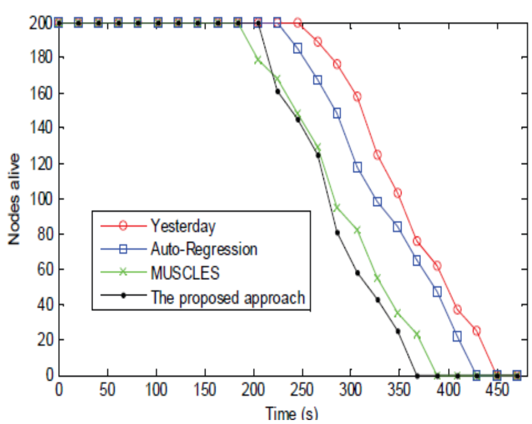

Figure1: Number of Nodes alive over time

VI. CONCLUSION AND FUTURE RESEARCH

In this paper, we propose a novel clustering algorithm for WSNs, which takes both the energy-efficiency problem and data correlations between the nodes into the unified consideration. The nodes can be assigned to a same cluster only when they have both close space distances and data correlations. Unfortunately, there are always some outliers that their readings are uncorrelated with most of their neighbors and it is impractical to generate some clusters for the outliers only. Therefore, after generating the clusters for most of the nodes, the outliers are assigned to the existing clusters based on the distances between the outliers and the cluster heads.

On one hand, close distances between the nodes make it energy-efficient for the nodes in a cluster to communicate with each other. On the other hand, high data correlations make it accurate for the cluster head to forecast the missing data. In addition, we design a distributed missing values forecasting approach to decrease data transmission in the network. Different from the traditional forecasting approaches, a pre-processing stage is integrated into the framework.

REFERENCES

[1] M. R. Palattella, M. Dohler, A. Grieco, G. Rizzo, J. Torsner, T. Engel, and L. Ladid, "Internet of things in the 5G era: Enablers, architecture, and business models," IEEE Journal on Selected Areas in Communications, vol. 34, pp. 510–527, Mar. 2016.

[2] A. Zanella, N. Bui, A. Castellani, and L. Vangelista, "Internet of things for smart cities," Internet of Things Journal IEEE, vol. 1, pp. 22–32, Feb. 2014.

[3] S. Cirani and M. Picone, "Wearable computing for the internet of things," It Professional, vol. 17, pp. 35–41, Sep. 2015.

[4] Mr. S.P. Santhoshkumar, Mr. C. B. Sivaparthipan, Dr. D. Prabakar, Dr. S Karthik, "Secure Encryption Technique with Keying Based Virtual Energy for Wireless Sensor Networks", International Journal of Advance Research in Computer Science and Management Studies, Volume 1, Issue 5, October 2013, http://www.ijarcsms.com/docs/paper/volume1/issue5/V1I5-0017.pdf.

[5] M. Simsek, A. Aijaz, M. Dohler, and J. Sachs, "5G-enabled tactile internet," IEEE Journal on Selected Areas in Communications, vol. 34, pp. 460–473, Mar. 2016.

[6] J. Fan, F. Han, and H. Liu, "Challenges of big data analysis," National Science Review, vol. 1, no. 2, pp. 293–314, 2014.

[7] Y. Shi, J. Zhang, and K. B. Letaief, “Group sparse beamforming for green Cloud-RAN,” IEEE Transactions on Wireless Communications, vol. 13, pp. 2809–2823, May 2014.

[8] M. Peng, S. Yan, K. Zhang, and C. Wang, “Fog-computing-based radio access networks: issues and challenges,” IEEE Network, vol. 30, pp. 46–53, Jul. 2016.

[9] Y. Shi, J. Zhang, K. B. Letaief, and B. Bai, “Large-scale convex optimization for ultra-dense Cloud-RAN,” IEEE Wireless Communications, vol. 22, pp. 84–91, Jun. 2015.

[10] M. Chiang and T. Zhang, “Fog and IoT: An overview of research opportunities,” IEEE Internet of Things Journal, vol. PP, no. 99, pp. 1–1, 2016.

[11] P. Derler, E. A. Lee, and A. S. Vincentelli, “Modeling cyber physical systems,” Proceedings of the IEEE, vol. 100, pp. 13–28, Jan. 2012.

[12] Y. Shi, J. Zhang, B. O’Donoghue, and K. B. Letaief, “Large-scale convex optimization for dense wireless cooperative networks,” IEEE Transactions on Signal Processing, vol. 63, pp. 4729–4743, Sep. 2015.

[13] K. Yang, Y. Shi, and Z. Ding, “Low-rank matrix completion for mobile edge caching in Fog-RAN via riemannian optimization,” in IEEE Global Communications Conference. (GOLBECOM),(Washington, DC, USA), 2016.