

An Introduction to Data Aggregation Techniques in Wireless Sensor Networks and Their Application

Prof. Rati. D. Joshi¹,

¹Asst.Prof. Deptt of Computer Science, Veerappa Nisty Engineering College, Shorapur ,India. 585224

Prof. Aruna. Rathod²

²Asst.Prof. Deptt of Computer Science Veerappa Nisty Engineering College, Shorapur, India.585224

Abstract-One data aggregation method in Wireless Sensor Networks (WSN) is sending local representative data to the sink node based on the spatial correlation of sampled data. In this paper, we highlight the problem that the recent spatial correlation model of sensor nodes's data is not appropriate for measuring the correlation in a complex environment. In addition, the representative data are inaccurate when compared with real data. Therefore, we propose the data density correlation degree, which is necessary to resolve this problem. The proposed correlation degree is a spatial correlation measurement that measures the correlation between a sensor node's data and its neighboring sensor node's data. Based on this correlation degree, a data density correlation degree (DDCD) clustering method is presented in detail so that the representative data have a low distortion on their correlated data in a WSN.

Index terms- Data aggregation, data density, correlation degree, wireless sensor networks, clustering methods.

I.INTRODUCTION

A WIRELESS sensor network (WSN) composed of self organized wireless sensor nodes distributed in a monitored area collects process and transmit data apart from the physical environment [1], [2]. The main goal of WSN is reliable detecting and accurately evaluating the events in the monitored area in the collected data for this purpose sensor nodes should be deployed densely. However, this will cause the overlapping of sensor nodes sensing area and special redundancy of adjacent sensor node data [3][4]. If every sensor node conveys collected data to the sink node the sensor node will consume much more energy. To reduce the amount of transmitted data in WSN, a greater number of correlation based data aggregation methods have been studied in the literature [5][11]. According to the level of sampled data in data aggregation strategy data aggregation method are grouped into three classes: data level aggregation, feature level aggregation, decision level aggregation [12]. Also based on the aggregation strategy we can divide the data level aggregation method into three types: in network type [5] [13], data compression type [6][14] and representative type[7][9][15][16]. It will take a long time to receive a reply from WSN in first type. In second type is of limited usefulness as it is too complex the

third type is sensitive to the correlation measurement of sensor nodes we focused on the representative type and try to find a proper correlation measurement of sensor nodes. We focused on the representative type and try to find a proper correlation measurement for sensor nodes. The major objective of the representative type is selecting a representative sensor node locally and sensing its observation to sink node therefore, the relative error between a representative data and its correlated data is a significant index for evaluating the represented performance. In the study of data aggregation method, the special correlation model between sensor nodes data is an important foundation that related to the accuracy of the aggregated data and the energy consumption of sensor nodes. Some researchers have systemically discussed spatial correlation models based on geographic locations of sensor nodes or statistic features of sensor nodes' data[7]-[9],[17][18]. The assumption of spatial correlation models based on sensor nodes location is the close sensor nodes are more correlated than the distant one. Thereby special correlation degree function is modeled to be non negative and decrease monotonically with the distance between sensor nodes. The sensor nodes are deployed in critical environments such as forest and sea areas. There is some distortion in sensing ability, due to some noise factor. The neighboring sensor nodes may not be correlated the spatial location of sensor nodes is not accurate hence it is difficult to accurately model the spatial correlation of sensor nodes. In our work a data density correlation degree (DDCD) was proposed to measure the spatial correlation of sampled data and try to resolve the drawbacks in existing spatial correlation models. DDCD clustering method was also presented in the WSN. In the DDCD clustering method sensor nodes which are in the same cluster have a high correlation degree and low distortion while those belonging to different clusters have low correlation degree. The remainder of this paper is organized as follows section II discusses related work on spatial correlation model in WSN. Section III presents the clustering method description. Section IV deals with relationship in [24]. At the network start up process a various applications of data aggregation. Section V presents the conclusion of the study.

II RELATED WORK

In present work on spatial correlation models are mainly based on the locations of sensor nodes' data. The spatial correlation model [7] simulated the transmitting process of data from source to the sink node. The spatial correlation between two sensor nodes is depicted by a function of the spatial distance between them. Four types of spatial correlation functions are given to capture the spatial temporal characteristics of point and field sources in WSN, the spatial temporal correlation models for point and field sources are theoretically analyzed in [20]. Meanwhile, the spatial temporal characteristics of point and field sources were analytically derived along with the distortion function. The correlation degree between two sensor nodes was obtained by the overlapping degree of their sensing areas [21]-[22]. This model is very convenient however it is difficult to pin point the location of sensor nodes' sensing areas change with their remaining energy. Thus, this type of spatial correlation is not accurate and impractical. In a practical environment, the area covered by a WSN is divided into some irregular parts. The sensor nodes in the same part have a high correlation in the data domain while those belonging to different parts have a low correlation. Along the boundary of two adjacent parts do not correlate. This practical situation is ignored in [7] [20] [22].

To resolve the drawbacks of spatial correlation models based on the spatial distance between the sensor nodes, the correlation of sensor nodes in the data domain was modeled in [8]. Unfortunately with the model proposed in [8] if two sensor nodes' data are the same at two different time intervals, the correlation degrees in these two time intervals will differ. The result doesn't agree with the reality. The definition of the spatial correlation weight considers the average spatial distance deviation between each sensor nodes' sampled data and that sampled by its neighbors within a predefined communication radius [18]. This spatial correlated weight is regarded as a measurement of the correlation degree. A semantic clustering architecture was proposed in [23] to group the sensor nodes according to semantic information and sensor nodes' connectivity properties. If a sensor node's sampled data satisfies the query sent from the sink node, it will select itself as a cluster head and start forming a cluster where all sensor nodes satisfy the same query. Therefore, the semantic correlated sensor nodes are those which satisfy the same query. This semantic correlation is appropriate for the data aggregation of in-network query type. In order to accurately detect the damage occurs gradually a semantic clustering model based on fuzzy system was proposed to find out the semantic neighborhood physical clustering is done to form a hierarchical physical organization consisted of two levels the upper level encompasses CHs and the lower consists of sensor nodes which are subordinated to one of the CHs when a sensor node's data satisfies a domain rule related to the event monitored by the WSN, this sensor node is

called the "candidate". If the data of the candidate changes the "candidate" becomes a semantic neighbor. Then, then CHs utilize the data of all the semantic neighbors which are in the same cluster are in the neighboring clusters to obtain an aggregation data by the fuzzy inference system as described in [24]. The semantic neighbors are correlated to the domain rules of the monitored event, so that this semantic clustering method is suitable for the event detection.

III. DESCRIPTION OF CLUSTERING METHOD

A. Data Density Correlation Degree

In a WSN, if a certain number of neighboring sensor nodes' data are close to a sensor node's data; this sensor node can represent its neighbors in the data domain. This representative sensor node is called the core sensor node. Definition 1: **Core sensor node.** Let us consider sensor node v has neighboring sensor nodes. They are respectively v_1, v_2, \dots, v_n . The data object of v is D . Its neighboring sensor nodes' data objects are respectively D_1, D_2, \dots, D_n . If there are N data object in D_1, D_2, \dots, D_n whose distances to D are a smaller amount than ϵ and $\min Pts \leq N \leq n$ then the sensor node v is called sensor node. Where $\min Pts$ is the amount threshold, ϵ is the data threshold. Instinctively, the larger the N is, the better representative the sensor node v is to those sensor nodes whose data objects are in ϵ -neighborhood of D . Meanwhile, high attention of the data objects in the ϵ -neighborhood of D implies that sensor node v has a high spatial correlation between it and these sensor nodes. Therefore, to determine the representation degree of v to those sensor nodes whose data objects are in ϵ -neighborhood of D in quantity, we proposed the data density correlation degree as shown in definition 2.

Definition 2: **Data density correlation degree.** Let sensor node v has n neighboring sensor nodes which are inside the cycle of the communication radius of v . They are v_1, v_2, \dots, v_n , respectively. The data object of v is D , and its neighboring sensor nodes' data respectively D_1, D_2, \dots, D_n . Among these n data objects, there are N data objects whose distance to D is not as much of than ϵ , and $\min Pts \leq N \leq n$. After that the data density correlation degree of sensor node v to the sensor node whose data objects are in ϵ -neighborhood of D is as follows.

Sim (v) =

$$0, N < \min pts$$

$$a_1 (1 - 1/\exp (N - \min pts)) + a_2 (1 - d\Delta/\epsilon) + a_3 (1 - d/\epsilon), N > \min pts \text{ ----- (1)}$$

Where $\min Pts$ is the amount threshold. ϵ is the data threshold. $d\Delta$ is the distance between D and the date centre of the data objects which are in the ϵ neighborhood of D . d is the average distance between the N data objects and D . $a_1 + a_2 + a_3 = 1$ If the data density correlation degree of sensor

node v is $Sim(v)$ defined by Eq.1, then we can obtain the properties of $Sim(v)$ as.

- 1) $Sim(v)$ increases with the increase of N , the amount of data objects which are in the ϵ -neighborhood of D ;
- 2) $Sim(v)$ increases with the increases with decreases of $d\Delta$, the distance between D and the data centre of the data objects which are in ϵ neighborhood of D ;
- 3) $Sim(v)$ increases with the decreases of d , the typical distance between D and those data objects which are in the ϵ neighborhood of D ;
- 4) $Sim(v) \in [0, 1]$. These properties are reliable with our intuitiveness. In definition 2, the date threshold ϵ guarantees that $Sim(v)$ will not be impacted by unconnected data. The amount threshold $min Pts$ is the minimum amount for sensor node v to represent some sensor nodes

IV APPLICATIONS OF DATA AGGREGATION

Based on the nature of the network, data aggregation can be done via data aggregation tree (DAT for flat networks) or by a clustering strategy for hierarchical networks. In flat network architecture, all nodes are equal and connections are set up between nodes that are within each other's radio range, although constrained by connectivity conditions and available resources. In a hierarchical network, all nodes typically function both as switches/routers, with one node in each cluster being designated as the cluster head (CH). The number of tiers within a hierarchical network can vary according to the number of sensor nodes. Traffic between nodes of different clusters must always be routed through their respective CHs or via gateway nodes that are responsible for maintaining connectivity among neighboring CHs. Although the hierarchical network architecture is energy efficient for collecting and aggregating data from the entire WSN or all nodes within a larger target region, using knowledge of their relative locations, flat network architecture is suitable for transferring data between source-destination pairs separated by a large number of hops. Data-diffusion is suitable for flat networks. A very common architecture used for data aggregation in flat networks is DAT. It begins with the route discovery phase when a designated node is selected as the root of the DAT and nodes in turn join the existing nodes of the DAT (also called parents) to construct the tree. A single network flow is assumed where a single data sink attempts to gather information from a number of data sources in the WSN. Data is collected from sources in proximity, aggregated at the parent nodes till it reaches the final aggregation point, i.e., the root node. With a DAT, a lower marginal energy cost is required in connecting additional sources to the sink along the shortest distance from the source to the DAT. Key Points in data aggregation are as follows: Nodes sense attributes over the entire network and route to nearby nodes. Node can receive different versions of same message from several neighboring nodes. Communication is usually performed in the aggregate. Neighboring nodes report similar data.

Combine data coming from different sources and routes to remove redundancy. There exists some research to study

the correlation in WSN. In [3], [4], [5], the theoretical aspects of the correlation are explored in depth. Basically, these studies aim to find the optimum rate to compress redundant information in the sensor observations. More recently, in [6], the relation between spatiotemporal bandwidth, distortion, and power for large WSNs has been investigated.

- **Spatial-temporal correlation**

In [7], Vuran et al. have proposed a theoretical framework to model the spatial and temporal correlations in WSN. Important key elements have been investigated to exploit the correlation in the WSN for the development of efficient communication protocols. In this paper, the authors show that the spatial-temporal correlation among the sensor observations is another significant and unique characteristic of the WSN which can be exploited to drastically enhance the overall network performance. Authors characterize the correlation in the WSN as follows:

- **Spatial correlation.**

Typical WSN applications require spatially dense sensor deployment in order to achieve satisfactory coverage. As a result, multiple sensors record information about a single event in the sensor field. Due to high density in the network topology, spatially proximal sensors' observations are highly correlated with the degree of correlation increasing with decreasing inter node separation.

- **Temporal correlation.**

Some of the WSN applications such as event tracking may require sensor nodes to periodically perform observation and transmission of the sensed event features. The nature of the energy-radiating physical phenomenon constitutes the temporal correlation between each consecutive observation of a sensor node. The degree of correlation between consecutive sensor measurements may vary according to the temporal variation characteristics of the phenomenon. In addition to the collaborative nature of the WSN, the existence of above mentioned spatial and temporal correlations bring significant potential advantages for the development of efficient communication protocols well-suited for the WSN paradigm. The main goal of the proposed framework in [7] is to enable the realization of efficient communication protocols that can exploit spatial and temporal correlations of the WSN paradigm. Based on the proposed framework, authors have discussed possible approaches to exploit these advantages intrinsic features of WSN for efficient medium access and reliable event transport.

V CONCLUSION

We have presented a data density correlation degree clustering (DDCD) method. The sensor nodes that have high correlation are divided into the same cluster, allowing more data aggregation data in clustering based data aggregation clustering method. Also, data conveyed to sink node can decrease. And also included the data aggregation technique to achieve the low data distortion in DDCD method.

REFERENCES

- [1] j.Yick, B.Mukharjee, and D ghosal, "wireless sensor network survey", *Compt. Netw.*, vol.52, no. 12,pp. 2292-2330, 2008.
- [2] L.M. Oliveria and J.J. Rodrigues, "Wireless sensor Networks: A survey on environmental monitoring," *J. Commun.*, vol. 6, no.2,pp 1796-2021,2011.
- [3] C.Zhu,C.Zheng, L. Shu, and G. Han, " A survey on coverage and connectivity issues in wireless sensor networks," *J. Netw. Comput. Appl.*, vol. 35, no 2,pp, 619-632,2012.
- [3] C. Zhu, C. Zheng, L. Shu, and G. Han, "A survey on coverage and connectivity issues in ireless sensor networks," *J. Netw. Compt. Appl. Vol. 35, no 2, pp. 619-632, 2012.*
- [4] G. Fan and S. Jin, "Coverage problem in wireless sensor network: A survey," *J. Netw.*, vol. 5, no.9, pp. 1033-1040,2010.
- [5]S. Madden, M. J. Franklin, J.M. Hellerstein, and W.Hong, "Tag a Tiny Aggregation service for ad-hoc sensor networks,"*ACM SIGOPS Operating syst. Rev.*, vol.36, no.9, no.1, pp. 131-146, 2002.
- [6]J. Zheng,P. Wong, and C. Li, "Distributed data aggregation using slepian- wolf coding in cluster-based wireless sensor networks", *IEEE Trans. Veh. Technol.*, vol.. 59, no..5, pp. 2564-2574, Jun. 2010.
- [7]M.C.Vuran ,O.B. Akan, and I.F.Akyildiz ,,"spatio-Temporal correlation:Theory and applications for wireless sensor networks," *compute.Netw.*, vol. 45, no.3, pp.245-259, 2004.
- [8]J.Yuan and H. Chen, "The optimized clustering technique based on spatial correlation in wireless sensor networks," in *Proc. IEEE youth Conf.inf., Comput. Telecommun. YC-ICT, Sep. 2008*, pp.411-414.
- [9]A.Rajewari and P.Kalaivaani, "Energy efficient routing protocol for wireless sendor networks using spatial correlation based medium access control protocol compared with IEEE 802.11," in *Proc. Int. Conf.PACC, Jul. 2011*, pp. 1-6.
- [10]J.N. Al-Karaki, R. Ul-Mustafa, and A.E. Kamal, " Data aggregation and routing in wireless sensor networks: Optimal and heuristic algorithms," *Compt. Netw.*, vol. 53, no. 7, pp. 945-960, 2009.
- [11] C. Hua and T. S. Yum, "Optimal routing and aggregation for maximizing lifetime of wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 16, no.4, pp. 892-903, Aug. 2008.
- [12] S.Iyengar, K. Chakrabarty, and H. Qi, "Introduction to special issue on 'distributed sensor networks for real-time systems with adaptive configuration'," *J. Franklin Inst.*, vol. 338, pp. 651-653, Jan. 2001.
- [13]S.Madden, R. Szewczyk, M.J Franklin, and D. Culler, " Supporting aggregate queries over ad-hoc wireless sensor networks," in *Proc. Mobile 4th IEEE Workshop Comput. Syst. Appl.*, Oct 2002, pp 49-58.
- [14]R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "On network correlated data gathering," in *Proc. IEEE Comput. Commun. Soc. 23rd Annu. Joint Conf. INFOCOM*, Mar. 2004, pp. 2571-2582.
- [15]M.C. Vuran and I.F. Akyildiz, "Spatial correlation-based collaborative medium access control in wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 14, no. 2, pp. 316-329, Apr. 2006.
- [16] G. A. Shah and M. Bozyigit, " exploiting energy aware spatial correlation in wireless sensor networks," in *Proc. 2nd Int. Conf. Commun. Syst. Sostw. MiddleWare, COMSWARE*, Jan. 2007, pp.1-6.
- [17]W.Guo, L. mZhai, L. Guo, and J. Shi, *Worm Propagation Control Based on Spatial Correlation in wireless Sensor Network*. Berlin, Germany: Springer-Verlag, 2012, pp. 68-77.
- [18] Y.Ma,Y.Guo, X. Tian, and M. Ghanem "Distributed clustering-based aggregation algorithm for spatial correlated sensor networks," *IEEE Sensors J.*, vol. 11, no. 3, pp 641-648, Mar. 2011.
- [19] C.Carvalho, D.G. Gomes, N. Agoulmine, and J.N. de Souza, " Improving prediction accuracy for WSN data reduction by applying multivariate Spatio-temporal correlation," *Sensors*, vol 11, no 11,pp. 10010-10037, 2011.
- [20] M.C. Vuran and O.B. Akan, "*Spatio-temporal characteristics of point and field sources in wireless sensor networks*," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2006, pp. 234-239.
- [21] N.Li,Y. Liu, F. Wu, and B.Tang, " WSN data distortion analysis and correlation model based on spatial locations," *J.Netw.*, vol.5, pp. 1442-1449, Dec. 2010.
- [22]R.K. Shakya Y.N.Singh, and N.K. Verma, "A novel spatial correlation model for wireless sensor network applications," in *Proc. 9th Int. Conf. WOCN*, Dec. 2012, pp. 1-6.
- [23]F.Bouhafs, M. Merabti, and H. Mokhtar, "A semantic clustering routing protocol for wireless sensor networks," in *Proc. 3rd IEEE Consum. Commun. Netw. Conf.*, Jan. 2006, pp. 351-355.
- [24]A. R. Rocha, L Pirmez, F. C. Delicato, E. Lemos, I. Santos, D.G. Gomes, et al., "WSNs clustering based on semantic neighborhood relationships," *Comput. Netw.*, vol. 56, no. 5, pp.1627-1645, 2012.
- [25]Institute for Nuclear Theory, Seattle, WA, USA. (2004). Intel Lab Data[Online]. Available: <http://db.csail.mit.edu/labdata/labdata.html>
- [26]Ecole polytechnique federale de Lausanne, Lausanne, Switzerland. (2006). LUCE [Online]. Available: <http://lcav.epfl.ch/cms/lang/en/pid/86035>