I.J. Wireless and Microwave Technologies, **2017, 2, 44-56**
Published Online March 2017 in MECS(http://www.mecs-press.net)
DOI: 10.5815/ijwmt.2017.02.05

Available online at **http://www.mecs-press.net/ijwmt**

Exploring an Effectiveness & Pitfalls of Correlational-based Data Aggregation Approaches in Sensor Network

Anand Gudnavar[a*], Dr. Rajashekhara[b]

[a]*Asst.Prof and Head, Dept. Of CSE, SCET Belagavi, Karnataka*
[b]*Head R&D, Venus Technologies, Bangalore, India*

Abstract

Data aggregation is one of the core processing in wireless sensor network which ensures that environmental data being captured reaches the user via base station. In order to ensure proper data aggregation, there are many underlying principles that need more attention as compared to more frequently visited routing and energy problems. We reviewed existing data aggregation schemes with special focus on data correlation scheme and found that there is still a large scope of investigation in this area. We find that there are only less number of research publications towards existing techniques of data aggregation using correlational-based approach. It was also explored that such techniques still does not focus much on data quality, computational complexity, inappropriate benchmarking, etc. This paper elaborates about all the unsolved issues which require dedicate focus of investigation towards enhancing the data reliability and data quality in aggregation process in wireless sensor network.

Index Terms: Data Aggregation, Data Correlation, Data Quality, Clustering, Wireless Sensor Network

1. Introduction

Wireless sensor network has already accomplished as global identity in research area associated with wireless communication. With an advantage of unattended data monitoring schemes, such network find much utility in the commercial market on wide range of applications e.g. habitat monitoring, healthcare, industrial monitoring etc. The topic is more than two decade old, but still it is associated with various forms of issues. At present, the research work toward sensor network are focused on addressing some of the significant issues viz. i) routing issues [1], ii) energy-consumption issues [2], iii) security issues [3], iv) optimization issues [4], v) load balancing issues [5], vi) traffic flow management / congestion issues [6] and many more. This where majority of the research works are absorbed into. However, we will like to highlight more on some basic problems

* Corresponding author. Tel.: 9886432683
E-mail address: anand_gudnavar@yahoo.co.in

which still remains much unaddressed i.e. data aggregation issues. Theoretical definition of data aggregation will mean a process by which a member node sense the raw data forward it to clusterhead and then clusterhead performs data fusion followed by forwarding the cumulative data to either sink directly (in case of single hop network) or to other clusterhead (in case of multihop network) [7]. Although, the theoretical concept of data aggregation sound so simple, but there are multiple problems associated with it. One simple problem can be stated as – what if two or more clusters sense the similar events but one slightly different time instances. In such case of multihop network, the aggregated data forwarded by clusterhead will never be able to find out that its neighboring aggregated data are also similar to a slight extent. This problem occurs due to difference in time stamping by the sensor as they normally work as per TDMA scheduling [8]. In such case, both the aggregated data (which are nearly similar) will be forwarded. The problem becomes much worst, if the scenario are either i) heterogeneous sensor network or ii) large scale network. This problem will lead to accumulation of lots of data in the base station which when forwarded to user finds more dependency of resource usage for performing analysis. Such aggregated data unwantedly increases machine and resource dependency, which is less addressed in any implementation studies till date. This problem is called as data correlation problem, which will mean how much the degree of correlation exists among the aggregated data. It is also known that applications of sensor network are basically of two types i) mission critical and i) time critical. Therefore, data correlation imposes bigger set of challenges in time critical application as it is not possible to ensure forwarding of non-redundant data on a stipulated amount of time on increasing round of data aggregation cycle. It is because increase of aggregation cycle will also means for a node to be nearing to dead condition sooner or later. Hence, at present, we don't have availability of any algorithms or techniques which has concretely addressed the problem of precise identification of data correlation during aggregation in wireless sensor network.

Therefore, the core goal of this paper is to perform an investigation of the existing techniques of data aggregation with more focus on reviewing effectiveness of existing data correlational schemes in wireless sensor network. Sub-Section 1.1 discusses about the background of the study that briefs about existing standards of research towards the data aggregation. Sub-Section 1.2 briefs about the problem identification followed by discussion of the proposed system in Section 1.3. Section 2 briefs above the essentials of data aggregation. We focus on less elaboration of theory as there are many study materials published in past discussing about theoretical background. Section 3 elaborates about existing techniques of data aggregation as well as data correlation with elaborated tabulation of its effectiveness and pitfalls. Section 4 discusses about research gap and finally paper summary is briefed in Section 5 as conclusion.

1.1. Background

This section discusses about the background of the study where the brief discussion is made for the existing techniques of data aggregation. We find that there are multiple research work [9][10], which have already briefed about data aggregation techniques in wireless sensor network. Therefore, we avoid repetitive discussion by highlighting only the less discussed points found in other studies. Fig.1 show the classification of data aggregation technique where specific focus is laid on to existing correlation-based techniques.

It can be seen that existing data aggregation techniques are classified into two types i.e. model-based aggregation and correlation-based aggregation. We find one effective techniques under model-based techniques i.e. PREMON or Predictive monitoring in wireless sensor network. Under correlation-based, we find three different techniques i.e. i) spatial, ii) temporal, and iii) Spatio-temporal schemes of data aggregation. Work carried out by Gupta et al. [11] named as EGCD is an example of spatial-correlation data usage whereas we find work carried out by Sharaf et al. [12] and Deshpande et al. [13] to be more of temporal approach. Sharaf et al. [12] have presented a scheme called TiNA whereas Deshpande et al. [13] introduced ARIMA. Finally, work carried out by Villas et al. [14] and Liu et al. [15] is a good example of spatiotemporal schemes focusing towards achieving energy efficiency.

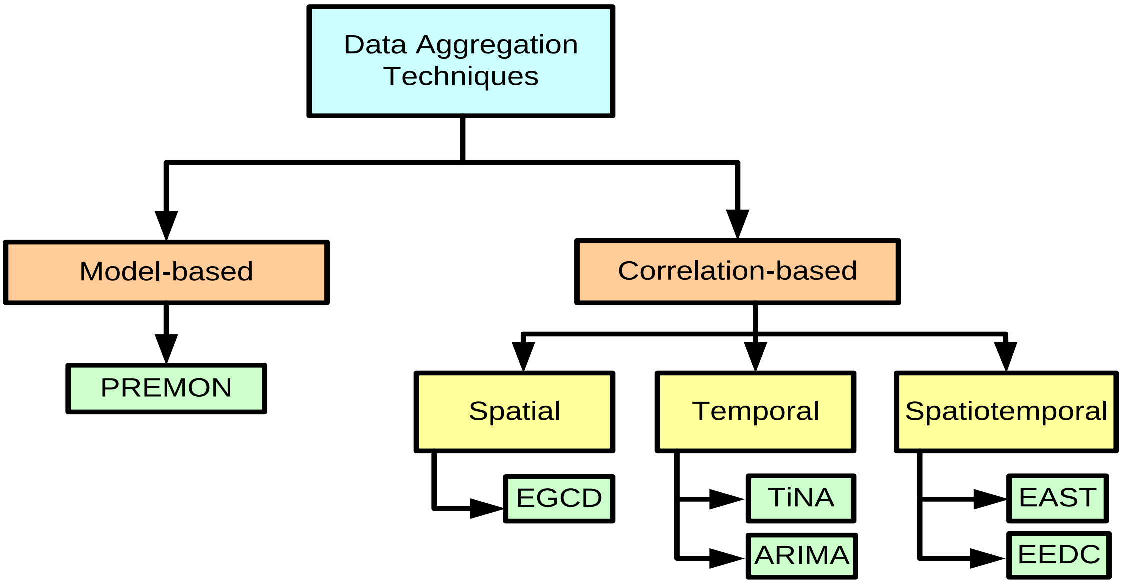

Fig.1. Type of data aggregation technique

1.2. The Problem

The problem identified in the existing system are as follow:

- Existing data aggregation techniques focuses less on data quality and inclined more on accomplishing communication targets.
- Existing schemes on correlation-based data aggregation has lesser extent of research work being done and hence the problem of effective data aggregation is still unsolved.
- Ineffective correlation analysis will lead to unwanted usage of resources which will lead to problems pertaining to routing and security mainly.

1.3. Proposed System

The previous section has discussed about the briefing of significant problems in the existing techniques of data aggregation in wireless sensor network. It was also spoken in Section 1.1 that there are already review papers existing at presents that has elaborated discussion of existing data aggregation techniques. However, we find that there are few review work focusing on effectiveness of correlational-based data aggregation techniques and discussing their beneficial factor and pitfalls. Therefore, in the proposed system, we review existing implementation techniques pertaining to data aggregation techniques and especially focusing on data correlation-based techniques. We review some of the standard papers published in last 5 years to investigate their contribution. The prime purpose of the proposed work discussed in this paper is to review existing techniques and finally discuss the research gap in this area.

2. Essentials of Data Correlation

Data correlation is basically a statistical technique to check the similarity of data processed by two sensors. Consider d_A and d_B be the data obtained by the sensor node A and B in wireless sensor network. Then, a mathematical equation is used to represent the condition of correlation as:

$$\frac{1}{\eta}\sum_{i=t_1}^{t_\eta}|C(d_A)-d_B|<e \quad (1)$$

In the above equation, the variable *C* is basically data correlation, *e* represents error, η represents number of sensors, t_i and t_η represents two time instances. Therefore, if the above equation is found to be satisfying than it is said that data of sensor *A* and sensor *B* at time t_i and t_η are correlated with each other at an error *e*. The significance of data correlation is as follows [16]:

- Data correlation is one of the important factors for selecting code rate from the sensory reading pertaining to source coding of distributed nature.
- Maximized value of data correlation confirms more generation of compressed data as well as minimized size of MAC payloads. These facts definitely enhance rate of data delivery.
- Event of retransmission of data packets can be minimized by extracting temporal correlational data.
- The data correlation can be also used to control the traffic congestion and thereby assists in formulating routes in stabilized links.

3. Studies on Existing Techniques

This section discusses about the existing technique that has been introduced during 2010-2016 pertaining to data dissemination techniques in wireless sensor network. For better effectiveness, we choose to consider only paper from research publishers with maximum impact factor.

3.1. Proposed System

Most recently, a very unique technique of improving the performance of data aggregation scheme was introduced by Uke and Thool [17]. The technique uses conventional theory of design pattern and Unified Modelling Language (UML) in order to analyze and improve data aggregation technique in wireless sensor network. According to author, such form of modelling assists in better visualization of communication pattern in sensor network during cycles of data aggregation. Consideration of cost parameter was seen in the work of Chen et al. [18], where the author have introduced algorithm for addressing the tradeoff in flows of local aggregator nodes and path with reduced cost leading to base station. The study outcome was evaluated with respect to differential entropy on increasing number of sources as well as cost and energy with increasing number of network size. Apart from cost, selection of clusterhead is another significant point in existing research work. Shwe and Chong [19] have introduced a linear coding approach for selection of clusterhead. The technique also introduced a multi-leveled architecture for assisting in data aggregation in sensor network. The study outcome is compared with the existing AODV (Adhoc On-Demand Distance Vector) routing protocol with respect to communication performance and delay. Koc and Korpeoglu [20] have carried out a study for enhancing the network lifetime in data aggregation considering multiple mobile sinks. The technique introduces both centralized and distributed aggregation technique with no dependency on pre-defined information of network. Sajedi and Saadati [21] have presented a graph-based data (minimum spanning tree) aggregation scheme for minimizing energy consumption during clustering. The outcome of the study was found to provide better energy efficiency with respect to LEACH algorithm. Nithyakalyani and Kumar [22] have presented a comparative study of clustering algorithm in sensor network. The authors have investigated

Voronoi schemes of clustering, fuzzy schemes, usage of evolution techniques, k-means clustering, etc. Apart from clustering schemes, scheduling schemes too enhance the performance of data aggregation to a large extent. Work on this direction was carried out by Chen et al. [23] where the authors have presented a scheduling approach based on duty cycle of stairs targeting to low-tier sensors. The technique also uses tree-based approach to design routing concept and uses the presented scheduling approach for controlling unwanted energy consumption and resist clock drift problems. Iabbassen and Moussaoui [24] have designed a technique that performs data aggregation using mobile agents in sensor network. The study outcome was found to be better in terms of reducing energy consumption.

Table 1. Summary of Existing Techniques of Data Aggregation

Author	**Problem**	**Technique**	**Advantage**	**Limitation**
Uke and Thool [17]	Modelling data aggregation	Unified Modelling Language	Involves design pattern	-No concrete analysis of outcomes
Chen et al. [18]	Cost minimization in data aggregation	Convex optimization	Highly effective for single sink	-No effective benchmarking -few analysis of communication performance
Shwe and Chong [19]	Clustering in multileveled network	Linear network coding	Reduced delay, increased packet delivery ratio	- No effective benchmarking - data quality not discussed
Koc and Korpeoglu [20]	Data aggregation on Multiple mobile sinks	Centralized and distributed data aggregation	Less message overhead	-no evidence of data reliability
Sajedi and Saadati [21]	Energy minimization during data aggregation	Minimum spanning tree based clustering	Better energy efficiency	-no evidence of data quality, overhead -complexity not analyzed
Chen et al. [23]	Scheduling of energy stages during idle states of data aggregation	Staircase scheduling algorithm	Energy efficient , higher reliability in communication	-data quality & redundancy management not discussed
Iabbassen and Moussaoui [24]	Reducing Energy/increasing scalability in Data aggregation	Mobile agents	Energy efficient	-Less evidence of numerical analysis -Complexity analysis not discussed - No effective benchmarking
Madhavi [25]	Energy consumption, inference	Distributed algorithm	Better control over energy	-Outcomes not benchmarked
Ren et al. [26]	Adoption of static routing in conventional data aggregation	Packet Attribute, dynamic routing	Supports both static and mobile aggregation schemes	-no study of complexity analysis -Doesn't address data quality
Yang et al. [27]	Extraction of sparse signals in sensor network	Compressive sensing	Estimates data quality using mean squared error,	-Doesn't use robust statistical model to support data quality (less scalability)
Lin et al. [28]	Security in data aggregation	Homomorphic encryption	-Effectively address secured data delivery	-Doesn't address data quality
Hoang et al. [29]	Optimization of energy utilization in data aggregation	Water drop algorithm	-highly energy efficient	-no study of complexity analysis -Doesn't address data quality
Holczer and Buttyan [30]	Security in aggregator node selection	Anonymous selection of aggregator node	-better security features (privacy)	-Doesn't address data quality
Jiang et al. [32]	Extracting statistical data during aggregation	Expectation maximization algorithm	Resistive towards node/link breakage	-no study of complexity analysis -No benchmarking
Ozdemir and Cam [33]	Securing data aggregation	Message authentication codes	-secured and communication effective	-No benchmarking

Madhavi [25] have presented a model for minimizing energy consumption and resisting interference in sensor network. The study outcome shows increasing latency over increasing nodes and transmission range. Ren et al. [26] have introduced an algorithm that considers time factor over the data packets while performing routing in sensor network. Significance attributed of data packets were used for enhancing the data aggregation performance. The author addresses a significant problem in data aggregation where it was stated that static routing is the frequent adoption in majority of the schemes of data aggregation which is incapable for forwarding data dynamically. The author addressed this scheme by introducing the concept of *packet attribute* which is data identifier to be sampled by multiple sensors. Another mean of enhancing the data aggregation performance was through compressed sensing approach. Yang et al. [27] have presented a study that extracts the sparse signals using concept of compressive sensing followed by signal reconstruction approach. The study outcome was evaluated with respect to mean squared error and data value on multiple observations. Data aggregation schemes also witnessed security-based approaches as performance enhancement techniques. Research in direction of incorporating secured data aggregation was carried out by Lin et al. [28] where the authors have used enhanced homomorphic encryption mechanism to incorporate secured mechanism. Hoang et al. [29] have presented an optimization techniques in data aggregation for achieving energy efficiency during data aggregation. The authors have presented *water-drop algorithm* which in targeted to improve the tree construction by maximizing the feasibility of aggregator node selection process. The study outcome was testified with respect to energy factor mainly. Holczer and Buttyan [30] have also presented a study to secure the process of aggregator node selection. There are also study by Gao et al. [31] that have joint usage of MIMO and conventional data aggregation for addressing the trade-off between increasing energy consumption and increasing data rate during data aggregation in sensor network. Jiang et al. [32] have presented a study for extracting statistical data using expectation maximization algorithm. Study by Ozdemir and Cam [33] have focused on secure data delivery during data aggregation in sensor network using message authentication codes. Table 1 summarizes the work done in enhancing performance of data aggregation in wireless sensor network.

3.2. Techniques Based On Enhancing Data Correlation Schemes

It was quite obvious from the previous section that very few approaches exists for data correlation scheme with respect to enhancing the data aggregation performance in wireless sensor network. Therefore, this section discusses about the techniques used for ensuring data correlational factor in sensor network.

Sun et al. [34] have presented a technique of data correlation in perspective of minimizing the energy dissipation in wireless sensor network. The author has emphasized that consideration of data correlation factor has positive effect in distributed topology of sensor network in controlling unwanted energy depletion from nodes. Relationship on energy efficiency and data correlation was also investigated by Vijayan and Raaza [35]. The author discussed the technique with respect to the uneven clustering mechanism over multi-hop networks. The study outcome shows reduction in message overhead, number of active clusters, and energy mainly. Studies towards correlational factor and massive data dissemination were carried out by Zhao et al. [36]. The author have presented a technique called as CoCo+ that constructs the structure of core correlation of a link in sensor network. It also have capability to address the overheads. Most recently, study on correlation was also carried out with respect to mobility of the targets. Zhu and Yu [37] have presented a technique that can identify the mobile target using correlational factor based on signal between the nodes with certain spatial distance. The outcome shows higher accuracy level of 99% approximately. Zou et al. [38] have presented a study to overcome the issue of conventional compressive sensing for data recovery in wireless sensor network. The presented technique allows integration of greedy approach over conventional compressive sensing that enhances the performance of data acquisition using spatial correlation. The study outcome was evaluated with respect to relative error, execution time, and energy parameters. Joshi et al. [39] have presented a study that uses both temporal and spatial correlation factor for enhancing the longevity of network lifetime. The technique presents a data correlational scheme along with clustering and active node scheduling to ensure energy efficiency. Similar direction of study considering both spatial and temporal data correlation was also seen in the work carried out by Kamal et al. [40]. The technique discussed was mainly intended for identifying outliers in sensory reading using fuzzy logic as well as an effective classification of suspicious observation from the reading. The methodology uses a sensor network dataset of only 23 real-sensors where the outcome was checked for outliers in terms of detection rate and false positive rate. Cui et al. [41] have presented an aggregation technique that uses spatial / temporal correlational data to identify the evolved data to be used in data aggregation and to avoid letting raw data in the data aggregation in sensor network. The study outcome for correlation was testified with respect to cosine similarity and root mean square error on TMH dataset.

Bertrand et al. [42] have introduced a study that perform analysis of correlational data of distributed nature to check its applicability for blind source separation. The presented technique performs analysis of correlated data among the signal sets over tree topology. The technique also allows estimation of significant direction from the observational data of sensors in order to extract respective sources. The significance of the correlation of data becomes quite high when it comes to extensive application of wireless sensor network for future e.g. Internet-of-Things (IoT). The study conducted by Bijarbooneh et al. [43] have used three input factors i.e. i) quality of link, ii) spatial correlational data, and iii) temporal correlational data. A constraint programming using greedy approach was used for developing an algorithm that can perform effective fusion of data applicable for upcoming IoT-based architectures. The study outcome was found to achieve better energy efficiency. Chidean et al. [44] have performed a study where spatial and temporal data were exploited in order to perform network partitioning for obtaining data coupling in sensor network. The authors have used multiple techniques e.g. principle component analysis, clustering with data coupling of second order, etc. The study outcome was tested for energy efficiency and signal-to-noise ratio and compared with LEACH. Das and Misra [45] have carried out a study of correlational data for investigating network management. The author have used cross layer approach and scheduling approach considering the input as correlated data. The prime target was to increase network lifetime. The technique also uses probability theory and learning algorithm for minimizing the energy depletion and delay in sensor network. Guo et al. [46] have presented a framework that can minimize cost of data aggregation in sensor network by permitting uploading of concurrent data from sensor to base station which are mobile. The technique also discusses about the travel period of mobile base station on multiple location considering the presence of spatial correlated data. The study outcome was found to achieve good performance in energy conservation. Similar direction of study was also carried out by Liu et al. [47]. The author presents a robust framework that performs extraction of knowledge from the correlational data present in neighbor sensor nodes and constructs a correlational graph for better routing analysis. The author uses simple descriptive statistics to derive the correlational decision technique and performing mining of it. The study outcome was evaluated with respect to energy consumption, mean delay, error and compared with conventional LEACH. Shakya [48] have presented a technique that uses fundamental spatial correlation function for modelling the characteristics of correlation corresponding to multiple events generated from multiple sensors using analytical and mathematical approach.

Pourpeighambar and Sabaei [49] have presented a study to address the energy consumption owing to moving targets in wireless sensor network. The author have used conventional rate-distortion theory as the base of the presented correlational model. The study outcome implemented with hybrid clustering technique is found to show better energy efficiency. Li and Wang [50] have presented a study using stochastic approach for conserving unwanted battery usage among sensors. The study uses statistical modelling of ARIMA (Auto-Regressive Integrated Moving Average) along with time-series analysis for evaluating the predicted data. The outcome of the study was testified with respect to error. Sun and Wu [51] have investigated sampling of multi-dimensional sensor network with both temporal and spatial correlated data. The authors claimed that adoption of space time sampling technique significantly minimize the error percentage of the correlated data computation in sensor network. The study outcome was testified with respect to error rate on both arbitrary and deterministic topology. Study of data correlation was also carried out by Lu et al. [52]. The technique presented have used clustering with greedy approach in order to sustain failures of correlations and issues of re-clustering in sensor network as a primary priority. The secondary priority of the presented technique also calls for performing data compression in order to enhance the network lifetime of wireless sensor network. The study outcome was evaluated with respect to network lifetime and ratio of data compression that was found to be better than existing algorithms.

Table 2 outlines the summary of the work carried out towards data correlation in wireless sensor network.

Table 2. Summary of Existing Techniques of Data Aggregation

Author	**Problem**	**Technique**	**Advantage**	**Limitation**
Sun et al. [34]	Energy minimization	Node hibernation	Highly controls energy depletion	-No benchmarking
Vijayan and Raaza [35]	Energy minimization	Uneven clustering	Energy efficient	-Not purely addressing data quality
Zhao et al. [36]	Minimize delay and enhance transmission	Node selection algorithm	Addresses link correlation	Doesn't address data correlation
Zhu and Yu [37]	Mobile target detection	Cross-correlation approach	Higher accuracy in detection	-applicable to less dynamic topology and small scale networks
Zou et al. [38]	Data acquisition	Spatial correlation, greedy approach, compressive sensing	Energy efficient	-no benchmarking
Joshi et al. [39]	Scheduling of active nodes, energy	Spatial and temporal correlation	Fault tolerance architecture	-No data analysis
Kamal et al. [40]	Outlier detection in sensor network	Fuzzy logic, spatial and temporal correlation	97% accuracy in outlier detection	-Analyzed on small network. -No benchmarking
Cui et al. [41]	Energy efficient data aggregation, data recovery	Spatial/temporal correlation	Highly reduced error in detection of evolved data	-doesn't measure complexity involved -communication performance not evaluated
Bertrand et al. [42]	Study of canonical correlation	Tree topology, canonical correlation (distributed)	Efficient convergence behaviour of algorithm	-Recursive nature leads to complexity -No benchmarking
Bijarbooneh et al. [43]	Data fusion for IoT	Constraint programming, greedy,	Energy efficient	-Analyzed on small network. -No benchmarking
Chidean et al. [44]	Self-organizing partition	Principle component analysis, clustering of coupled data	Good Energy retention	-doesn't measure complexity involved in algorithm processing
Das and Misra [45]	Energy efficiency	Cross layer approach, probability theory, learning algorithms	Enhance network lifetime, accurate data collection, and minimal delay	-doesn't measure complexity involved in algorithm processing
Guo et al. [46]	Optimizing time for data uploading	Framework for uploading concurrent data	Reduced delay, reduced cost, and reduced energy	-doesn't measure complexity involved in algorithm processing
Liu et al. [47]	Maintain uniformity in correlational data	Correlational framework, mining, descriptive statistics	Reduced delay/error/energy consumption	-doesn't measure complexity involved in algorithm processing
Shakya [48]	Minimizing degree of spatial correlation in dense network	Spatial correlation	Low redundancy, low energy expense, maximum throughput	-doesn't measure complexity involved in algorithm processing
Pourpeighambar and Sabaei [49]	Energy consumption for mobile data aggregation	Rate distortion theory, hybrid clustering	Minimize error with increasing target velocity	-No benchmarking -doesn't measure complexity involved in algorithm processing
Li and Wang [50]	Prediction of correlated data	Time-series, ARIMA	Minimize error	-No benchmarking -doesn't measure complexity involved in algorithm processing
Sun and Wu [51]	Space time sampling	Mathematical modelling of sampling on multiple dimension	Reduced error occurrences	-No benchmarking -doesn't measure complexity involved in algorithm processing
Lu et al. [52]	Correlational failures, re-clustering	Clustering using data correlation	Increased network lifetime	-doesn't measure complexity involved in algorithm processing

4. Research Gap

This section discusses about the research gap being explored after reviewing the existing techniques towards data aggregation techniques. The viewpoint of the research gaps are mainly to explore a cost effective algorithm/technique/modelling to ensure better data quality from data aggregation in wireless sensor network. Following are the outlines of the research gaps:

- **Few studies towards data correlation**: We find that there are only existence of less than 18% of IEEE Transactions paper discussed about data correlational techniques in wireless sensor network published between 2010-1016. This proportion of research paper in data correlation is extremely less numbered as compared to research papers on other problems in sensor network e.g. routing problem, energy problems, security problem, load-balancing problem etc. Hence, there is a need to increase more investigation towards this topic.
- **Less focus on data quality**: Out of above mentioned numbers of published research papers, 15% of the papers have introduced the concept of data correlation for solving energy problems as majority followed by clustering techniques. Therefore, the focus on data quality is highly ignored in existing system. There are abundant number of research work focusing on data fusion, aggregation, mathematical modelling, but very few have proved that their model enhances the data quality in every cycle of data aggregation.
- **Less involvement of extensive performance study**: 99% of the work towards existing data correlational modelling has majorly focused on energy parameters. However, along with energy there are many other parameters e.g. throughput, delay, latency, packet delivery ratio, bandwidth consumption, impact of multiple traffic load, impact of multiple mobility factor, impact of dense/sparse network etc. Such parameters were never seen to be even prioritized in existing study.
- **Lack of Complexity Evaluation**: It is said that a normal MicaZ mote or TelosB mote can only use 48 kilobyte of physical memory. A closer look into the outcome analysis of the existing system will show that none of the existing studies till date were able to prove their study compatibility with real-time motes memory usage. Although, there are mixed study approach with both hardware and software. But tradeoff exists in many ways viz. hardware based approaches have used less number of motes and study outcomes accomplished in very controlled environment. Hence, this outcomes are definitely not scalable in nature. On the other hand, computational model doesn’t focus on convergence test or show the performance of time/space complexity for any of research papers published till date. Therefore, the outcomes are not supported with enough evidence of compatibility with real-time sensor operation.
- **Less effective benchmarking**: Benchmarking makes the claims of the researchers more concrete to be accepted. However, existing study outcomes are found to be showcased with multiple graphical trends which doesn’t bear some potential resembles with the study. The existing research shows the graphical outcomes mainly exhibiting the individual results and very less on comparative analysis with other standard techniques. Although, some of the researchers have used performance comparative evaluation with other standard techniques, but it is quite backdated. For an example, majority of the study uses LEACH as comparative analysis parameter for energy efficiency, whereas in reality PEGASIS performs better than LEACH, but there is not such work on energy efficiency using data correlation that has used PEGASIS, TEEN/APTEEN, HEED, etc. just for an example. More extent of benchmarking guides the research work more accurately in future direction.

5. Conclusions

This paper has discussed the existing mechanism of data aggregation with special stress on studying data correlation-based approaches. After reviewing the existing techniques, we find that majority of the existing techniques on data aggregation are focused on achieving energy efficiency, which is a very positive fact.

However, the limitations is ignorance of involvement of low and high-level statistical computation just to ensure data correlation from multiple clustering. The existing studies are found to focus less on data quality under which circumstance the reliability score is quite less on such data. In this scenario, it invites other associated problems too e.g. security and routing problems. Studies focusing on data correlation-based approaches are found with lesser extent of benchmarking which raise questions of acceptance of such techniques from global research viewpoint. However, all the existing mechanism are limited by a scope and further research can be only investigated if such scopes are identified and amended to further increased scope. Hence, our future direction of study will be to develop a novel model of data correlation-based approach of aggregation that can ensure better data quality and goes in line with communication parameters of wireless sensor network.

References

[1] S.P. Singh and S.C. Sharma, “A survey on cluster based routing protocols in wireless sensor networks” Elsevier Procedia Computer Science, Vol.45, pp.687-695, 2015

[2] A. Boonsongsrikul, S. Kocijancic and S. Suppharangsan, "Effective energy consumption on wireless sensor networks: Survey and challenges," Information & Communication Technology Electronics & Microelectronics (MIPRO), 36th International Convention, Opatija, pp. 469-473.2013

[3] A. Boonsongsrikul, S. Kocijancic and S. Suppharangsan, "Effective energy consumption on wireless sensor networks: Survey and challenges," Information & Communication Technology Electronics & Microelectronics (MIPRO), 2013 36th International Convention, pp. 469-473, 2013.

[4] M.A. Adnan, M.A. Razzaque, I. Ahmed, and I.F. Isnin, I.F, “ Bio-Mimic optimization strategies in wireless sensor networks: a survey”. Sensors, Vol. 14(1), pp.299-345,2013.

[5] R.K. Ghotekar, and D.C. Mehetre, “Load Balancing for Achieving the Network Lifetime in WSN-A Survey”, International Journal of Innovative Research in Advanced Engineering (IJIRAE), Vol. 1, Issue. 4, 2014

[6] E. Dashkova and A. Gurtov, "Survey on Congestion Control Mechanisms for Wireless Sensor Networks", Springer, Volume 7469, pp. 75-85, 2012

[7] S. Rani, S.H. Ahmed, "Multi-hop Routing in Wireless Sensor Networks: An Overview, Taxonomy, and Research Challenges", Springer, Technology & Engineering, pp. 69, 2015

[8] I.M. M. E. Emary, S. Ramakrishnan, "Wireless Sensor Networks: From Theory to Applications", CRC Press, Computers, pp. 799, 2013

[9] H. Dhasian, and P. Balasubramanian, "Survey of data aggregation techniques using soft computing in wireless sensor networks", Information Security, IET, Vol.7 (4), pp.336-342,2013

[10] S. Sirsikar, and S. Anavatti, “Issues of Data Aggregation Methods in Wireless Sensor Network: A Survey”, Elsevier Procedia Computer Science, Vol. 49, pp.194-201, 2015.

[11] H. Gupta, H., V., Das, S.R., V. Chowdhary, “Efficient gathering of correlated data in sensor networks”, In: MobiHoc, Urbana-Champaign, Illinois, USA, 2005

[12] M.A. Sharaf, J. Beaver, A. Labrinidis, P.K. Chrysanthis, “TiNA: a scheme for temporal coherency-aware in-network aggregation”, In: MobiDE’03, San Diego, California, USA, 2003

[13] A. Deshpande, C. Guestrin, J.M. Hellerstein, S.R. Madden, W. Hong, “Model-driven data acquisition in sensor networks”, In: 30th VLDB Conference, Toronto, Canada, 2004

[14] L.A. Villas, A. Boukerche, D.L. Guidoni, H.A.B.F, de Oliveir, R.B. de Araujo, A.A.F. Loureiro, “An energy-aware spatio-temporal correlation mechanism to perform efficient data collection in wireless sensor networks”, Comput. Commun, Vol.36, pp. 1056–1066, 2013

[15] C. Liu, W. Kui, J. Pei, “An energy-efficient data collection framework for wireless sensor networks by exploiting spatiotemporal correlation”, IEEE Trans. Parallel Distrib Syst, Vol. 18(7), 2007

[16] H. Jin, “Handbook of Research on Developments and Trends in Wireless Sensor Networks: From Principle to Practice: From Principle to Practice”, IGI Global, pp. 556, 2010

[17] S. Ukea, R. Thoolb, "UML Based Modeling for Data Aggregation in Secured Wireless Sensor Network", Elsevier,International Conference on Information Security & Privacy (ICISP2015), Vol. 78,pp. 706-713, 2015,

[18] C. Chen, K. Lee, J-S. Park, and S. J. Baek, "Research Article Minimum Cost Data Aggregation for Wireless Sensor Networks Computing Functions of Sensed Data", Hindawi, Journal of Sensors,Article ID 506909, pp. 17, 2015

[19] H.Y. Shwe, and P. H. J. Chong, “Building Efficient Multi-level Wireless Sensor Networks with Clustering”, Springer International Publishing In Wireless Internet, pp. 8-13, 2014.

[20] M. Koç, & I. Korpeoglu, “Coordinated movement of multiple mobile sinks in a wireless sensor network for improved lifetime”, EURASIP Journal on Wireless Communications and Networking, (1), 1-16, 2015

[21] H. Sajedi, and Z. Saadati, “A Hybrid Structure for Data Aggregation in Wireless Sensor Network”, Journal of Computational Engineering, 2014.

[22] S. Nithyakalyani, and S. S. Kumar, “Data aggregation in wireless sensor network using node clustering algorithms—a comparative study”, In Information & Communication Technologies (ICT), 2013 IEEE Conference, pp. 508-513, 2013.

[23] J. Chen, Y. Wang, Y. Wang, and C. Hu, “Stair Scheduling for Data Collection in Wireless Sensor Networks”, International Journal of Distributed Sensor Networks, 2013.

[24] D. Iabbassen, and S. Moussaoui, “Mobile line based data dissemination protocol for wireless sensor networks”, Journal of Emerging Technologies in Web Intelligence, Vol. 5(1), pp.4-11, 2013

[25] S. Madhavi, “Dynamic Distributed Power Adaption for Data Aggregation in Ubiquitous Sensor Networks”, Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology, Vol. 3(2), pp.47-60, 2013

[26] F. Ren, J. Zhang, Y. Wu, T. He, C. Chen, and C. Lin, “Attribute-aware data aggregation using potential-based dynamic routing in wireless sensor networks”, Parallel and Distributed Systems, IEEE Transactions, Vol. 24, No. 5, pp.881-892, 2013

[27] G. Yang, M. Xiao, and S. Zhang, "Data aggregation scheme based on compressed sensing in wireless sensor network”, Springer Berlin Heidelberg, In Information Computing and Applications, pp. 556-561, 2012.

[28] Y.H. Lin, S.Y. Chang, and H.M. Sun, “CDAMA: concealed data aggregation scheme for multiple applications in wireless sensor networks”, .Knowledge and Data Engineering, IEEE Transactions, Vol.25 (7), pp.1471-1483, 2013.

[29] D.C. Hoang, R. Kumar, and S.K. Panda, “Optimal data aggregation tree in wireless sensor networks based on intelligent water drops algorithm”, Wireless Sensor Systems, IET, Vol. 2(3), pp.282-292, 2012.

[30] T. Holczer, and L. Buttyán, “Anonymous aggregator election and data aggregation in wireless sensor networks”,. International Journal of Distributed Sensor Networks, 2011.

[31] Q. Gao, Y. Zuo, J. Zhang, and X-H. Peng, “Improving energy efficiency in a wireless sensor network by combining cooperative MIMO with data aggregation”, Vehicular Technology, IEEE Transactions, Vol.59, No. 8, pp. 3956-3965, 2010

[32] H. Jiang, S. Jin, and C. Wang, “Parameter-based data aggregation for statistical information extraction in wireless sensor networks”, Vehicular Technology, IEEE Transactions, Vol. 59(8), pp.3992-4001, 2010

[33] S. Ozdemir, and H. Çam, “Integration of false data detection with data aggregation and confidential transmission in wireless sensor networks”, .IEEE/ACM Transactions on Networking (TON), Vol.18 (3), pp.736-749, 2010

[34] S-J. Sun, C-j. Zhao, P. Guo, and X-q. Li, “The Research of Reducing Energy Consumption Method on WSN Based on the Data Correlation”, In Proceedings of the 6th International Asia Conference on Industrial Engineering and Management Innovation, pp. 3-10, 2016.

[35] K. Vijayan, and A. Raaza, “A novel cluster arrangement energy efficient routing protocol for Wireless Sensor Networks”, Indian Journal of Science and Technology, Vol. 9, No. 2, 2016

[36] Z. Zhao, J. Bu, W. Dong, T. Gu, and X. Xu, “CoCo+: Exploiting correlated core for energy efficient dissemination in wireless sensor networks”, Ad Hoc Networks, Vol.37, pp.404-417, 2016.

[37] H. Zhu, and F.Q. Yu, “A cross-correlation technique for vehicle detections in wireless magnetic sensor network”, 2016

[38] Z-q.Zou, Z-t. Li, S.Shen, and R-c.Wang, “Research Article Energy-Efficient Data Recovery via Greedy Algorithm for Wireless Sensor Networks”, Hindawi Publishing Corporation, pp. 9, 2016

[39] K. Joshi, S. Siddiqui, and S. Fatima, “Data-Based Correlation Scheme”, In Proceedings of the International Conference on Recent Cognizance in Wireless Communication & Image Processing, pp. 297-304, 2016.

[40] S. Kamal, R.A. Ramadan, and E.R. Fawzy, E.R., “Smart outlier detection of wireless sensor network”, Facta Universitatis, Series: Electronics and Energetics, Vol.29 (3), pp.383-393, 2015

[41] J. Cui, and F. Valois, “Simba: Similar-evolution based aggregation in Wireless Sensor Networks”, In Wireless Days, 2016.

[42] A. Bertrand, and M. Moonen, “Distributed canonical correlation analysis in wireless sensor networks with application to distributed blind source separation”, Signal Processing, IEEE Transactions, Vol.63, No. 18, pp. 4800-4813, 2015.

[43] F. Bijarbooneh, W. Du, E. Ngai, X. Fu, and J. Liu, “Cloud-assisted data fusion and sensor selection for internet-of-things”, 2014

[44] M.I. Chidean, E. Morgado, E. d. Arco, J. R-Bargueno, and A. J. Caamano, “Scalable Data-Coupled Clustering for Large Scale WSN”, Wireless Communications, IEEE Transactions, Vol. 14, No. 9, pp. 4681-4694, 2015.

[45] S.N. Das, and S. Misra, “Correlation-aware cross-layer design for network management of wireless sensor networks”, Wireless Sensor Systems, IET, Vol.5 (6), pp.263-270, 2015.

[46] S. Guo, Y. Yang, and C. Wang, C., “DaGCM: A concurrent data uploading framework for mobile data gathering in wireless sensor networks”, 2015

[47] C. Liu, J. Luo, and Y. Song, “Correlation-Model-Based Data Aggregation in wireless sensor networks”, In Fuzzy Systems and Knowledge Discovery (FSKD), 2015 12th International Conference, pp. 822-827, 2015

[48] R.K. Shakya, "Spatial Correlation based Efficient communication Protocols for Wireless Sensor Networks", IIT, Kanpur, 2014

[49] S.B. Pourpeighambar, and M. Sabaei, M, “Spatial correlation aware protocols for efficient data aggregation of moving object in wireless sensor networks”, Scientia Iranica, Vol. 20(3), pp.695-709, 2013

[50] G. Li, and Y. Wang, “Automatic ARIMA modeling-based data aggregation scheme in wireless sensor networks”, EURASIP Journal on Wireless Communications and Networking, (1), pp.1-13, 2013.

[51] N. Sun, and J. Wu, “Optimum sampling in spatial-temporally correlated wireless sensor networks”, EURASIP Journal on Wireless Communications and Networking, No. 1, pp. 1-18, 2013.

[52] J. Lu, B. Zhang, and L. Xu, “A data correlation-based wireless sensor network clustering algorithm”, In Computer Application and System Modeling (ICCASM), International Conference, Vol. 8, pp. V8-61, 2010.

Authors' Profiles

Anand Gudnavar is working as Assistant Professor and Head, in the Department of Computer Science and Engineering at Shaikh College of Engineering and Technology Belagavi, India. He received his M.Tech. and B.E. degrees from Visvesvaraya Technological University in 2006 and 2003. He has thirteen years of Academic and Teaching experience. Currently he is pursuing research in the area of WSN data aggregation.

Dr. Rajashekhara received Ph.D. from IIT Bombay and M.Tech from University of Mysore. He has more than eighteen years of experience in Industry, research and teaching at reputed organisations. His research and product development experience includes Communications, Digital signal and image processing, Machine learning, Pattern recognition, Biometric authentication and analysis, System modelling and design, Currently he is heading R&D division in Venus Technologies, Bangalore. Karnataka, India.

How to cite this paper: Anand Gudnavar, Rajashekhara,"Exploring an Effectiveness & Pitfalls of Correlational-based Data Aggregation Approaches in Sensor Network", International Journal of Wireless and Microwave Technologies(IJWMT), Vol.7, No.2, pp.44-56, 2017.DOI: 10.5815/ijwmt.2017.02.05