Ternary Hierarchical Clustering with Data De-Duplication in Sensor Networks

Sreeram Indraneel[1] and Kurra Raja Sekhara Rao[2]
1. Associate Professor, CSE, St.Anns Engineering College,Chirala,A.P.,India.email:Sreeram.indraneel@gmail.com
2. Professor, CSE, K.L.University,Guntur-522502.India. email:krr@kluniversity.in

Abstract

Wireless Sensor Networks are resource constrained application specific adhoc networks. All the previous works have concentrated on the power aware techniques such as data aggregation, coordination of node operations for effective routes and query processing etc. All these techniques have helped to collect and communicate the field data of numerous applications at the cost of degrading the quality of the data. This tradeoff between resource consumption and the quality of information flow will in turn decrease the utility of sensed data. In this paper, we proposed a new approach for gathering and communicating the sensed data with the help of clustering and De-Duplication technique. We have explained the significance of de-duplication technique by considering the multi-media data transmission in wireless sensor networks.

Keywords: Ad-hoc networks, Clustering, De-duplication, Multi-Media Data, etc.

1. INTRODUCTION

In Prevailing days, Quality improvement is important for any application of sensor networks. One of major problems in sensor network data is identified as the presence of redundancy in the sensed data. Till date, the research in various dimensions in sensor networks concentrates on low power consumption in the forms of processing, transmission and reception of data etc. By the very nature of sensor networks lot of duplicated data will be generated by the sensor nodes deployed in the application field area. In this paper, we concentrate on improving the Quality of the transmitted data before actual transmission to the longer distances. Instead of transmitting all the sensed data to the end user which leads to more power consumption because of transmission and reception of lot of message packets containing the sensed data, we can decrease the no. of transmissions and receptions by de-duplicating the data at the gateways hierarchically. However, for accuracy we need the repeated data. But at the same time we have to be aware of the power consumption because of power constrained nature of wireless sensor networks. So, we have concentrated on decreasing the redundancy of the sensed data by applying various de-duplication techniques to optimize the sensed data without losing the generality such that the quality of the transmitted data from the application field to the end user can be improved.

Earlier works on these lines were proposed by various authors, which are related to mechanisms for information flow management in sensor networks. In sensor networks, nodes are the devices with limited battery life, storage capacity, and processing capabilities. Various techniques such as in-network data aggregation to preserve resources, and/or the use of coordination of node operations for effective routing and query processing are published. In-network aggregation is not using de-duplication. It concentrates on evaluating the aggregate statistics from data (e.g., sum, average, etc). Therefore, studies on these lines focus on the design of data-aggregation strategies [1, 2, 3, and 4] as well as associated routing of queries and query results [5] in order to allow for reduced communication costs. Here we note that there is an implicit tradeoff between resource consumption and the quality of information flow. An aggregate is effectively an approximation of the information. Aggressive aggregation leads to reduction in the quality and hence utility of information flow. Coordinated communication of sensory data is the recent work focused on minimizing the communication cost required for the evaluation of multi-predicate queries scheduled across network nodes [5]. In this work, a dynamic-programming algorithm is used to determine an optimal execution order of queries. Examples of other studies that proposed coordination strategies include techniques that ensure efficient storage, caching, and exchange of spatio-temporal data [6, 7] or aggregates thereof [8]. In our work, we focus on reducing overheads by adding Data De-duplication functionality to a subset of nodes, without concern to resource or energy constraints.

The rest of this paper is organized as follows. Section 2 deals with the basic idea of De-Duplication. In Section 3 we deal with the Deduplication in Wireless Sensor Networks using Ternary Hierarchical Clustering with data deduplication procedure. Section 4 deals with the Benefits of De-Duplication in wireless sensor networks. Section 5 deals with Multi-media data transmission in wireless sensor networks. Section 6 deals with the conclusion and future perspective.

2. DATA DE-DUPLICATION

It is a specific data compression technique which eliminates duplicate copies of redundant data. This is also called as intelligent data compression. This technique can be applied to network data transfers to reduce the number of bytes that must be sent. In the de-duplication process, unique chunks of data, or byte patterns, are identified and compared to the stored copy and whenever a match occurs, the redundant chunk is replaced with a small reference that points to the stored chunk. De-duplication may occur in two ways namely called as Post-Process De-duplication, In-line De-duplication.

2.1. Post-process De-duplication

In this method, new data is first stored on the storage device and then analyzes the data for finding duplicates. Here there is no need to wait for the hash calculations and lookup to be completed before storing the data, which ensures that store performance is not degraded. But the drawback with this method is, it store duplicate data for a short time which is an issue if the storage capacity of the system is very small.

2.2. In-line De-duplication

In this process the deduplication hash calculations are created on the target device as the data enters the device in real time. If the device finds a block that is already stored on the system, it does not store the new block and simply references to the existing block. The benefit of in-line deduplication over post-process deduplication is that it requires less storage as data is not duplicated. Here the drawback is, hash calculations and lookups takes long time which means that the data storage into the device can be slower thereby reducing the backup throughput of the device.

3. DE-DUPLICATION IN WIRELESS SENSOR NETWORKS USING TERNARY HIERARCHICAL CLUSTERING:

In wireless sensor networks we can implement the De-duplication technique to greatly enhance delivery efficiency of messages from the sensor nodes deployed at the application field area to the base station or sink node or the end user of the application data. In wireless sensor networks the nodes are shared across geographically and administratively distributed systems, and use more heterogeneous hardware. So we can use the three level hierarchy constituting member nodes, cluster heads, gateways. De-duplication is an approach for processing distributed data across huge datasets using a large number of nodes. It can take advantage of locality of data generated in wireless sensor networks by forwarding aggregated data in the cluster heads to the pre determined nearby gateway nodes choosed for implementing data De-duplication to decrease the amount of transmitted data.

Basic Procedure:

Step 1.We consider some of the highly powered nodes which are equally distributed across the sensor field are assigned as gate way nodes responsible for De-duplication.

Step 2. The Gateway nodes will be sending beacon frames indicating their presence. The neighboring cluster heads will send the acceptance as member cluster head for the nearby gateway node.

Step 3. Prepare the input data set for gateways–The member cluster head node is assigned the unique key value for all the data associated with it.

Step 4. Apply the Hash code on the generated data sets with corresponding assigned key values.

Step 5. Send the generated hash value and data pairs to De-duplicating cluster heads.

Step 6. Run the user-provided De-duplication technique on the data collected from member cluster heads.

Step 7. Produce the final output by sorting based on the hash values of the outcome data.

4. BENEFITS OF DE-DUPLICATION IN WIRELESS SENSOR NETWORKS:

The benefits of using De-duplication techniques in wireless sensor networks are four fold. Firstly, because the wireless sensor nodes are with less memory space, the data De-duplication will conserve the overall memory space of the total nodes deployed in the field area.

Secondly, the wireless sensor networks are power constrained where most of the overall network power is consumed for transmission and reception of messages. By data De-duplication no of average transmitted messages throughout the network which includes transmissions between peers at various hierarchies of the network and transmissions from sensor field to sink node will be decreased which in turn decreases the power consumptions.

Third, because the bandwidth is also resource constrained in sensor networks, De-duplication will help in effective utilization of bandwidth by eliminating the duplicates predominantly.

Fourth, with De-duplication the network traffic predominantly decreases and consequently data loss due to collisions can be decreased.

5. MULTI-MEDIA DATA TRANSMISSION IN WIRELESS SENSOR NETWORKS:

The miniaturization in electronics contributed to the availability of multimedia equipment with low cost such as CMOS cameras and microphones which lead to the development of wireless multimedia sensor networks (WMSN). Here the nodes are able to retrieve multimedia content such as the audio, video, still images, along with data from environmental sensors. Wireless multimedia sensor networks will expand the application horizon of wireless sensor networks to the new fields such as battle field surveillance, home automation monitoring networks multimedia, storage of potentially activities, control systems traffic, medical surveillance environmental monitoring, location services, and industrial process control. Traditionally the data generated by sensor network will be of small size consisting of sensory information and communication control information. With the advent of WMSN large volumes of multimedia data is generated. In general multimedia data is highly redundant which needs the techniques such as De-duplication which are generally applied on Big data in internet servers, cloud computing etc. Here we have proposed the application of this technique at the third level hierarchical nodes called gateways in ternary cluster hierarchy to decrease the transmission overhead between gateways and the sink node or the end user of application data.

6 CONCLUSION AND FUTURE PERSPECTIVE

With the increase of utilization of wireless sensor networks for huge data generating applications such as Multimedia wireless sensor networks (MWSNs), lot of redundancy in the application data is noticed. Since the sensor networks are power constrained networks, there is a need for effective utilization of resources. So we have proposed the application of Data De-duplication technique at Gateway nodes where huge data will be collected from the neighboring cluster heads. This implementation will predominantly decrease the transmission power consumption and overhead between Gateway node and end user of the application. The future perspective will be identifying the novel techniques for positioning the gateway nodes at appropriate positions in the network to decrease the communication over head between cluster heads and gateways.

REFERENCES

[1] K. Dasgupta, K. Kalpakis, and P. Namjoshi. An efficient clustering-based heuristic for data gathering and aggregation in sensor networks. In in Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC, pages 1948–1953, 2003.

[2] A.Goel and D. Estrin. Simultaneous optimization for concave costs: Single sink aggregation or single source buy-at-bulk. In In Proc. of the 14 th Symposium on Discrete Algorithms (SODA, pages 499–505, 2003.

[3] J.Heidemann, F. Silva, C. Intanagonwiwat, R. Govindan, D. Estrin, and D. Ganesan. Building efficient wireless sensor networks with low-level naming, 2001.

[4] H. Luo, Y. Liu, and S. K. Das. Distributed algorithm for en route aggregation decision in wireless sensor networks. IEEE Transactions on Mobile Computing, 8, January 2009.

[5] G.Chatzimilioudis, H. Hakkoymaz, N. Mamoulis, and D. Gunopulos. Operator placement for snapshot multi-predicate queries in wireless sensor networks. In Proceedings of the 2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware, MDM '09, pages 21–30. IEEE Computer Society, 2009.

[6] H.Morcos, A. Bestavros, and A. Matta. Amorphous Placement and Informed Diffusion for Timely Field Monitoring by Autonomous, Resource-Constrained, Mobile Sensors. In Proceedings of SECON'08: The IEEE Conference on Sensor, Mesh and Ad Hoc communications and Networks, pages 469–477, San Francisco, CA, June 2008.

[7] H. Morcos, A. Bestavros, and I. Matta. Preferential Field Coverage Through Detour-Based Mobility Coordination. In Proceedings of Med-Hoc-Net'10: The IFIP/IEEE Mediterranean Ad Hoc Networking Workshop, Jun-Les-Pins, France, June 2010.

[8] H. Morcos, G. Atia, A. Bestavros, and A. Matta. An Information Theoretic Framework for Field Monitoring Using Autonomously Mobile Sensors. In Proceedings of DCOSS'08: The 4th IEEE/ACM International Conference on Distributed Computing in Sensor Systems, Santorini, Greece, June 2008.

[9] Vivek Katiyar, Narottam Chand, Surender Soni A Survey on Clustering Algorithms for Heterogeneous Wireless Sensor Network Int. J. Advanced Networking and Applications Volume: 02, Issue: 04, Pages: 745-754 (2011).

[10] Azer Bestavros, Dóra ErdOs, VatcheIshakian, Andrei Lapets, Evimaria Terzi The Filter-Placement Problem and its Application to Content De-Duplication Computer Science Department, Boston University Boston, MA 02215, USA.

BIBLIOGRAPHY OF AUTHORS

	Mr. SREERAM.INDRANEEL received the M.E degree in CSE from Sathyabama University, Chennai and he is doing his Ph.D in Acharya Nagarjuna University ,Guntur in the area of Computer Networks under the guidance of Dr.KURRA.RAJASEKHARA RAO. He is currently working as an Associate Professor in the department of Computer Science and Engineering in St.Ann's Engineering college, Chirala. His research interests are in Computer Networks, Sensor Networks.
	Dr.KURRA.RAJASEKHARA RAO is currently working as Professor in Computer Science and Engineering Department i & Dean-Administration, K L University, Guntur. His research interests are software reliability.