

Data De-Duplication Aware Clustering for Optimized Storage in Wireless Multimedia Sensor Networks

Sreeram Indraneel¹ and Kurra Raja Sekhara Rao²

¹Associate Professor, CSE, St. Anns Engineering College, Chirala, A.P., India.
email: Sreeram.indraneel@gmail.com

²Professor, CSE, K.L. University, Guntur-522502. India.
krr@kluniversity.in

Abstract

The progressive growth in the scope and scale of wireless sensor networks have explored new dimensions of potential research in this domain. Traditionally they are viewed as resource constrained and application specific localized adhoc networks. The capabilities of the sensor node have increased many fold with the advent of new sensors, miniaturization of electronic devices and storage facilities. Now a days the sensor nodes are equipped with on chip CMOS cameras, Flash memory and considerably large storage space and power resources. Many changes in communication protocols have led to the possible integration of heterogeneous sensor networks into the existing infrastructure based networks. With all these capabilities sensor nodes are capable to transmit multimedia data such as video, audio, image etc. seamlessly over the large scale networks leading to large scale Wireless Multimedia Sensor Networks (WMSNS). For optimizing the data transmission and storage of the generated huge multimedia data we have proposed De-Duplication aware clustering in WMSNS.

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

1. Introduction

Traditional wireless sensor networks are viewed as small scale localized adhoc networks. The sensor nodes are capable of generating the scalar data such as values of temperature, pressure, humidity etc in the field area. With the miniaturization of electronics, availability of on chip CMOS cameras have made the sensor nodes capable of recording video, images also. The nature of video and image data needs more storage space, bandwidth when compared to the normal wireless sensor networks. These properties of newly evolved sensor networks are have transformed wireless sensor networks to wireless multimedia sensor networks(WMSNS). One of major problems in WMSNS is identified as the presence of redundancy in the sensed data. 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.

The rest of this paper is organized as follows. Section 2 discusses the related work in the area. Section 3 deals with WMSNS. Section 4 deals with the basic idea of De-Duplication. In Section 5 we describe various phases involved in the Deduplication aware

Clustering applied for WMSNS. Section 6 deals with experimental results using synthesized video data section 7 is conclusion and future perspective.

2. Related Work

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 at various hierarchies, without concern to resource or energy constraints.

3. Wireless Multimedia Sensor Networks

Wireless multimedia sensor network (WMSN) consists of WMS nodes, gateway nodes with storage, sink nodes, and so on. Then, a typical WMSN architecture is proposed in [12], as shown in Figure 1. WMSN is widely used in many areas including visual monitoring, individual positioning, industrial control, intelligent transportation, environmental monitoring, smart home, and telemedicine [11–13].

Multimedia sensor nodes have limited resources such as energy capacity of battery, storage space, and computing power, while the information they acquire features large data size and high requirement for computing.

The data in-network processing technology [14-15], as one of the key technologies used to save bandwidth and network energy in data-centered WMSN, can eliminate the redundancy of source data and minimize the data traffic between nodes by applying data fusion technology or data compression. The data fusion technology uses different routing methods to combine data packages and eliminate redundant data from them. This technology relies on three basic modules routing algorithm, data fusion, and data presentation [15]. However, in addition to the advantage of high data aggregation, the application of data fusion technology may also lose raw data structure, while data compression can prevent this problem. Data compression has been applied to WMSN, and the compression algorithm can be divided into two categories distributed data compression scheme and local data compression scheme.

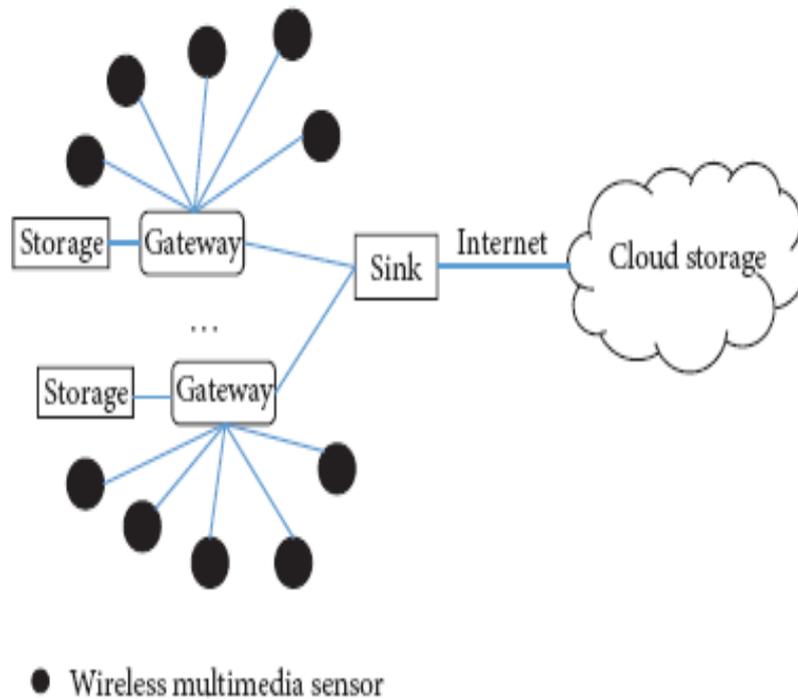


FIGURE 1: A typical system architecture of WMSN.

4. 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.

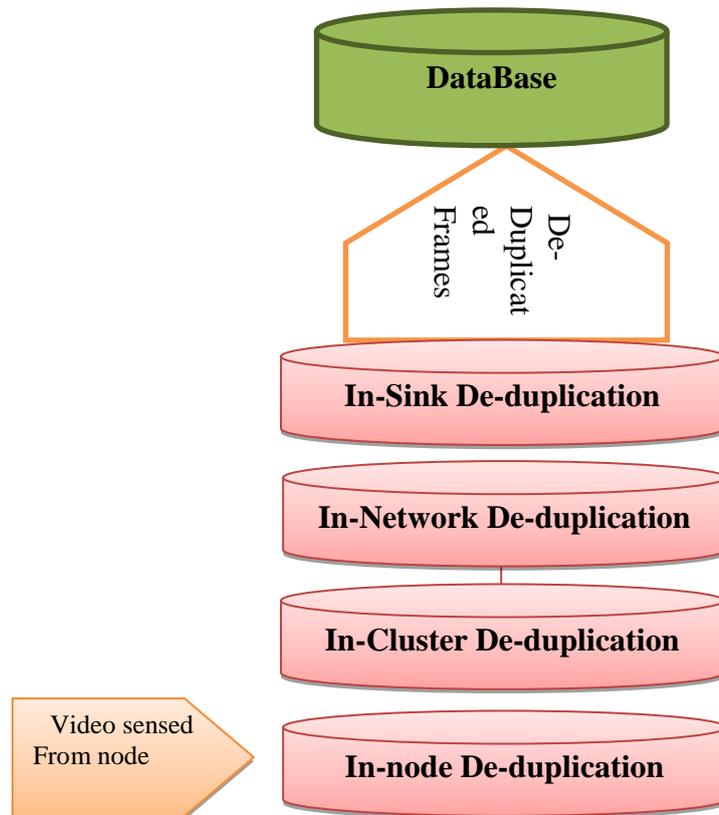
4.1. Post-process De-duplication

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.

4.2. 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.

5. Data De-Duplication Aware Clustering in WMSNS



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 predetermined nearby gateway nodes choosed for implementing data De-duplication to decrease the amount of transmitted data.

5.1. Basic Procedure:

Phase 1. The adhoc sensor network is configured with video sensing nodes. Some of the nodes having more power and processing capabilities distributed across the network. The nodes in the network are grouped into clusters by considering highly powered nodes as clustered heads.

Phase 2: The member nodes which are equipped with video sensors will record the video and convert the video into consecutive frames as shown in Figure 3. The background of the image which is same for all the consecutive frames of the video are subtracted using background subtraction method in the video sensing node. The background image is shown in Figure 4. This is quoted as In-node De-duplication.

Phase 3: The image data set from video sensor nodes are collected at the cluster head and processed for similarity matching. If the match is found only one image is transmitted and

reference index is forwarded for rest of the similar images. This is quoted as In-Cluster De-duplication.

Phase 4: The same process is repeated for the datasets transmitted from multiple cluster heads to the Gateways. This is quoted as In-network De-duplication.

Phase 5: Finally the redundant data is removed at the Sink node of the deployed video sensor network.

6. Experimental Results using Synthesized Video Data:



Figure 3. Conversion of Video into Consecutive Image Frames

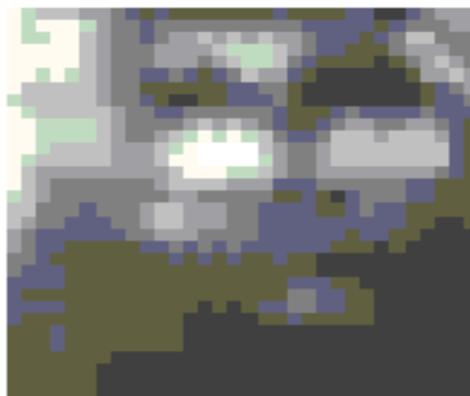


Figure 4. Background Image Subtracted from each Frame of above Sample Dataset



Figure 4. Before De-duplication



Figure 5. After De-duplication

7. Conclusion

In our work we have proposed the application of Data De-duplication technique at various hierarchies of a WMSN. This implementation will predominantly decrease the

transmission power consumption and overhead between Gateway node and end user of the application. We have illustrated our work using the sample recorded video which is sampled at appropriate sampling rate and generated a sequence of image frames. To detect the moving object in the video, the generated sequence of frames is subjected to the back ground suppression. As part of this experiment, back ground image is detected as shown in experimental results. The frames are generated using Homogeneous based Poisson process. The generated frames are verified for redundancy using De-duplication technique and the data without duplicates is generated to transmit over the network. The future perspective will be identifying the novel techniques for positioning the gateway nodes at appropriate positions in the network to decrease the communication overhead 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", Proceedings of the IEEE Wireless Communications and Networking Conference, WCNC, (2003), pp. 1948–1953.
- [2] A. Goel and D. Estrin, "Simultaneous optimization for concave costs: Single sink aggregation or single source buy-at-bulk", Proc. of the 14th Symposium on Discrete Algorithms (SODA), (2003), pp. 499–505.
- [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, (2009) January 8.
- [5] G. Chatzimilioudis, H. Hakkoymaz, N. Mamoulis and D. Gunopulos, "Operator placement for snapshot multi-predicate queries in wireless sensor networks", Proceedings of the 2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware, MDM '09, IEEE Computer Society, (2009), pp. 21–30.
- [6] H. Morcos, A. Bestavros, and A. Matta, "Amorphous Placement and Informed Diffusion for Timely Field Monitoring by Autonomous, Resource-Constrained, Mobile Sensors", Proceedings of SECON'08: The IEEE Conference on Sensor, Mesh and Ad Hoc communications and Networks, San Francisco, CA, (2008) June, pp. 469-477.
- [7] H. Morcos, A. Bestavros and I. Matta, "Preferential Field Coverage Through Detour-Based Mobility Coordination", Proceedings of Med-Hoc-Net'10: The IFIP/IEEE Mediterranean Ad Hoc Networking Workshop, Jun-Les-Pins, France, (2010) June.
- [8] H. Morcos, G. Atia, A. Bestavros and A. Matta, "An Information Theoretic Framework for Field Monitoring Using Autonomously Mobile Sensors", Proceedings of DCOSS'08: The 4th IEEE/ACM International Conference on Distributed Computing in Sensor Systems, Santorini, Greece, (2008) June.
- [9] V. Katiyar, N. Chand and S. Soni, "A Survey on Clustering Algorithms for Heterogeneous Wireless Sensor Network", Advanced Networking and Applications, vol. 02, no. 04, (2011), pp. 745-754.
- [10] A. Bestavros, D. ErdOs, V. Ishakian, A. Lapets and E. Terzi, "The Filter-Placement Problem and its Application to Content De-Duplication Computer Science Department", Boston University Boston, MA 02215, USA.
- [11] M. Tubaishat and S. Madria, "Sensor networks: an overview", IEEE Potentials, vol. 22, no. 2, (2003), pp. 20-23.
- [12] I. F. Akyildiz, T. Melodia and K. R. Chowdhury, "A survey on wireless multimedia sensor networks", Computer Networks, vol. 51, no. 4, (2007), pp. 921–960.
- [13] F. Hu and S. Kumar, "QoS considerations in wireless sensor networks for telemedicine", Internet Multimedia Management Systems IV, Proceedings of SPIE, vol. 5242, (2003) November, pp. 217–227.
- [14] J. N. Al-Karaki, R. UI-Mustafa and A. E. Kamal, "Data aggregation and routing in wireless sensor networks: optimal and heuristic algorithms", Computer Networks, vol. 53, no. 7, (2009), pp. 945–960.
- [15] Z. Lu, Y. Wen, R. Fan, S. Tan and J. Biswas, "Toward efficient distributed algorithms for in-network binary operator tree placement in wireless sensor networks", IEEE Journal on Selected Areas in Communications, vol. 31, no. 4, (2013), pp. 743–755.

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 previously worked as Professor in Computer Science and Engineering Department in K L University, Guntur. He is currently working as Director, Sri prakash College of Engineering, Tuni, Andhra Pradesh, India.. He has more than 20 years of teaching experience both in Under Graduate and Post Graduate Programs. He has published several papers in different National and International Journals. His areas of research interest are Software Engineering, Agile Techniques, Reliability Studies, Quality Control methods, wireless sensor networks.