Weekly Report (2009-10-25)

Hongxing Li

1. Current Problems

1) For energy-optimization problem, current results are mainly based on a matrix with arbitrary weights of channel gain. We need to define the problem with the SINR constraints and confirm the channel gain for each signal. Is it still proportional to $1/d^\alpha$ or any combined effect of two concurrent signals?
2) For data aggregation with successive interference cancellation, is it applicable for sensor nodes to implement successive interference cancellation? If not, is it possible for sensor nodes to just relay the signal to the sink and let sink to realize interference cancellation?
   If combined optimization of both energy-consumption and scheduling latency cannot be tackled, is it reasonable to aim on the achievable throughput over per watt, which is the unit of energy consumed per second.
3) Different from distributive (*min, max, sum, count*) and algebraic (*plus, minus, average, variance*) aggregation functions, which can be easily implemented with *convergecast* operations over a *breadth first search* tree, the holistic (*median, k^{th} smallest, or largest value*) functions are more difficult to solve by current aggregation algorithms with constraint communication complexity. Although there is some work conducted under the graph model, which will be introduced in *paper study* section, we have not found any work under the SINR model. It can be one of our research problems to analyze the scheduling complexity and design efficient deterministic or approximate algorithms.

2. Paper Study

- [1] proposes a distributed approximation algorithm (Q-Digest) for quantile queries. The Q-Digest algorithm builds a so-called q-digest tree on each sensor based on their local information. A merging and compression operation will be executed on the intermediate nodes when local q-digest trees are aggregated to the sink. As a result, the sink will get a merged q-digest tree which can approximate the data distribution in network. Analysis shows that the relative error can be bounded nicely and will be improved with larger message size.
- [2] explores the median value problem and its general case: k-order statistics problem. Different from previous works that require linear communication and storage complexity for both issues, this paper presents both deterministic and approximate algorithms with ploylog and polyloglog communication complexity respectively. The idea behind this improvement is that multiple pass of aggregation is allowed over the network and binary search is utilized. The performance on communication complexity can be even better if prefix-checking is implemented for approximate algorithm.
- Different from the theoretical style of [2], [3] is more like a system work. A practical order statistics service, which is composed of three parts - initialization, validation and update, is implemented. Only initialization part requires full picture of the reading values from all sensor nodes. While in validation and update parts, sensors just periodically feed back constant-sized information indicating the changes of reading values. Similar with [2], binary search is applied in the update part to enhance the performance.
- [4] targets the k-selection problem in wireless sensor networks. Different from commonly held opinion that convergecast operation is not feasible for k-selection problem, the paper presents both randomized and deterministic distributed algorithms for k-selection problem utilizing convergecasting. The basic idea is to iteratively process the candidate set and reduce its size. The analysis demonstrates that the time complexity can be bounded as $O(D\log_D n)$ and $O(D\log_D^2 n)$ for randomized and deterministic algorithms respectively. Here, D is the network diameter and n is the number of nodes.
- Although [5] has not been published, we still find it a meaningful work in data collection, aggregation and selection. It addresses on the complexity of these functions from three different angles: time complexity, message complexity and energy complexity. Corresponding lower bounds for optimal methods and upper bounds are presented. Efficient distributed algorithms are also introduced for each functionality. The result in this paper serves as good guidance for further works under graph model. We may explore our bounds for those applications under the SINR model.

References

[1] N. Shrivastava, C. Buragohain, D. Agrawal and S. Suri, *Medians and Beyond: New Aggregation Techniques for Sensor Networks*, In proceedings of SenSys'04, Nov. 3-5, 2004, Baltimore, Maryland, USA.

[2] B. Patt-Shamir, *A Note on Efficient Aggregate Queries in Sensor Networks*, In Theoretical Computer Science, 370 (2007), pp. 254-264.

[3] L.P. Cox, M. Castro and A. Rowstron, *POS: A Practical Order Statistics Service for Wireless Sensor Networks*, In proceedings of ICDCS'06, July 2006, Lisbon, Portugal.

[4] F. Kuhn, T. Locher and R. Wattenhofer, *Distributed Selection: A Missing Piece of Data Aggregation*, Communications of the ACM, Vol. 51, NO. 9, Sept. 2008.

[5] X.-Y. Li, Y. Wang and Y. Wang, *Complexity of Data Collection, Aggregation, and Selection for Wireless Sensor Networks*, under submittion.