Ant-aggregation: Ant Colony Algorithm for optimal data aggregation in Wireless Sensor Networks

Rajiv Misra, and Chittaranjan Mandal

***Abstract*—Data aggregation is an essential paradigm for energy efficient routing in energy constraint wireless sensor networks. The complexity of optimal data aggregation is NP-hard. Ant colony system, a population-based algorithm, provides natural and intrinsic way of exploration of search space in optimization settings in determining optimal data aggregation. The simulation results shows improvement in energy efficiency depends on number of source nodes in sensor network which is 45% energy efficiency using optimal aggregation compared to approximate aggregation schemes in moderate number of source whereas 20% energy efficiency in large number of source nodes. The proposed Ant-Aggregation algorithm is simulated in MATLAB.**

***Index Terms*—Wireless sensor networks, data aggregation, Ant Colony System.**

I. INTRODUCTION

Recent developments in MEMS have made these tiny sensor nodes available in large numbers, to be used in a wide range of applications: military, environmental monitoring, etc. A wireless sensor network operates in an unattended environment, with limited computational and sensing capabilities capable of sensing, computing and wirelessly communicating [1]. In order to effectively utilize wireless sensor nodes, we need to minimize energy consumption in the design of sensor network protocols and algorithms. Since the sensor nodes have irreplaceable, batteries with limited power capacity, it is essential that the network be energy efficient in order to maximize the life span of the network. Large number of sensor nodes have to be networked together, direct transmissions from any specified node to a distant base station is not used, as sensor nodes that are farther away from the base station will have their power sources drained much faster than those nodes that are closer to the base station. On the other hand, minimum energy multi-hop routing scheme will result in rapidly drain energy resources of the nodes, since these nodes engage in the forwarding of a large number of data messages (on behalf of other nodes) to the base station. Thus solution is to use multi-hop communication with in-network aggregation of correlated data. In this, sensor nodes collect, processes, and forward the data from all the sensor nodes to BS. The application of a aggregation approach helps reduce the amount of information that needs to be transmitted by performing data fusion at the aggregate points before forwarding the data to the end user.

The rest of the paper is organized as follows: In next section application for which this work is relevant is discussed. Section 3 briefly touches upon related prior work. Section 4 presents the problem formulation and the contribution of this paper in bird's eye view. Following sections 5, 6 &7 discuss the ant-aggregation algorithm using ACO, implementation, results and conclusion

Rajiv Misra, School of Information Technology, Indian Institute of Technology, Kharagpur(India)721302 email: rajivm@iitkgp.ernet.in
Chittaranjan Mandal, Computer Science & Engg, Indian Instiute of Technology, Kharagpur(India)721302 email: chitta@iitkp.ac.in

1-4244-0340-5/06/$20.00 ©2006 IEEE

II. DATA AGGREGATION

The application model considered for this work consists of a single destination (basestation) and multiple sources. Since the nodes are wirelessly connected which communicates to neighbors in vicinity, therefore multi-hop communication is used to reach the destination. It is assumed that density of nodes gives a connected node graph. For a application setting, data aggregation is applied in the network.

"Data aggregation merges message data in-network while traversing through network" it is also termed as data fusion. The aggregation gain can be measured as (original – aggregated)/original in the given application message size. The aggregation suffers from delay termed as aggregation delay. There is a tradeoff in delay and gain in aggregation. The simulation study reveals that energy-efficiency is related to number of source nodes in correlated sensing.

III. RELATED WORK

In-network data aggregation is an important in energy constraint sensor network which exploits correlated sensing data and aggregates at the intermediate nodes reducing the number of messages exchanged network. In data gathering application large amount of communication is reduced by in-network aggregation achieving maximum lifetime of network. Optimal aggregation tree problem is NP-Hard[5] which is equivalent to Steiner tree[1],weighted set cover[2] problems. Approximation algorithms for finding optimal aggregation are Greedy Incremental Tree(GIT)[1],Shortest Paths Tree(SPT),Center at Nearest Source(CNS). Active research in area of sensor network aims for finding efficient approximation algorithms for optimal aggregation problem. Optimal aggregation is modeled as combinatorial optimization problem which is solved using population based metaheuristic approach Ant Colony Optimization (ACO). The ACO has been used in prior works for routing in sensor network [4]. This is first attempt to the author's knowledge of applying ACO for optimal aggregation problem in sensor network. The proposed Ant-aggregation determines optimal aggregation in network using ant colony system heuristics. The behavior of ants modeled as artificial ants, is natural to use to solve this combinatorial optimization problem.

IV. THE PROBLEM & CONTRIBUTION

A. *The Problem*

To determine optimal in-network data aggregation points in sensor network. The optimal aggregation tree in sensor network is shown to be NP-Hard in [1,2] due to combinatorial search space. Swarm intelligence based technique is used in optimal data aggregation problem.

B. *Our Contribution*

1. The problem of determining optimal data aggregation is modeled as Ant system optimization.
2. Ant-aggregation algorithm, constructs iteratively aggregation tree in network which converges to an optimal (minimum) cost solution. The Simulation is carried out in MATLAB.
3. Optimal aggregation is compared, with opportunistic aggregation [1] and greedy incremental aggregation [2] algorithms.

Simulation results show that energy-efficiency in data-aggregation depends on number of source nodes in sensor network which is up to 45% for moderate size source nodes and up to 20% for high number of source nodes in sensor networks

V. ANT SYSTEM &OPTIMIZATION

In ACO(Ant colony optimization) a colony of artificial ants is used to construct solutions guided by the pheromone trails and heuristic information. ACO was inspired by the foraging behavior of real ants. This behavior enables ants to find shortest paths between food sources and their nest. Initially, ants explore the area surrounding their nest in a random manner. As soon as an ant finds a source of food, it evaluates quantity and quality of the food and carries some of this food to the nest. During the return trip, the ant deposits a pheromone trail on the ground. The quantity of pheromone deposited, which may depend on the quantity and quality of the food, will guide other ants to the food source. The indirect communication between the ants via the pheromone trails allows them to find the shortest path between their nest and food sources. This functionality of real ant colonies is exploited in artificial ant colonies in order to solve Optimization problems.

In ACO algorithms the pheromone trails are simulated via a parameterized probabilistic model that is called the pheromone model. The pheromone model consists of a set of model parameters whose values are called the pheromone values. The basic ingredient of ACO algorithm is a constructive heuristic that is used for probabilistically constructing solutions using the pheromone values.

In general, the ACO approach attempts to solve a CO problem by iterating the following two steps (1) Solutions are constructed using a pheromone model, that is, a parameterized probability distribution over the solution space. (2) The solutions that were constructed in earlier iterations are used to modify the pheromone values in a way that is deemed to bias the search toward high quality solutions.

Ant Colony Algorithm
Input: weighted graph, neighborhood info. **While** termination not met **do** Compute-initial pheromone, node dist potential Schedule activities Ant based solution construction Pheromone update Node distance potential update **End** activities Best <-best solution in population of solution
Output: Best, candidate to optimal solution

VI. ACO FOR DATA AGGREGATION

A. *Ant Colony Optimization for Optimal Aggregation Tree*

The Algorithm is runs in two passes. In forward of the algorithm, the route is constructed by one of the ants in which other ants search the nearest point of previous discovered route. The points where multiple ants join are aggregation nodes. In the backward pass nodes of the discovered path are given weight in form of node potential which indicates heuristics for reaching to destination(for the first ant or other nodes) or nearest aggregation point(for other ants) and pheromone trails is the heuristics to communicate other ants of the route discovered. Ants tries to follow the route to get pheromone eventually converges to the optimal route. Non-optimal route pheromone gets evaporated with time. The aggregation points on the optimal tree identify data aggregation. The indicator in data aggregation points gives estimate of number of paths aggregates in it.

B. *Heuristics for global optimization of aggregation tree*

In large sensor network, finding optimal aggregation tree is NP-Hard problem. The reduction is weighted set covering problem. In the Ant-aggregation algorithm, given a permutation of source nodes constructs aggregation tree associated with a cost which is the local best aggregation tree. The algorithm iterates to search the global best and the convergence of algorithm gives the optimal aggregation tree from combinatorial space. Thus, the best aggregation tree constructed by ant routing in iterations is remembered. Further, giving early aggregation more weight in cost function will converge in optimal aggregation points.

C. *Data Correlation*

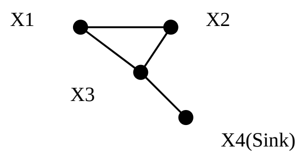

Fig.1: Correlation

The aggregated traffic in sensor network is local traffic generated by node when sensing, and the transit from neighbors. In wireless sensor networks, data collected at neighboring nodes are often correlated, due to large redundancy in sensed information by multiple sensor nodes. In order to reduce communication, the redundant information can be removed without any loss by capturing correlation between nodes. For example in the above fig.1 (i)If x1 and x2 use node x3 as relay(ii) x1 uses x2 as relay and x2 transmits aggregate to node x3.

Let denote X_j the random variable measured at node j. Nodes without side information coded data with $H(X_j)=R$ bits. But, at intermediate nodes on aggregation tree have side information from child nodes resulting in reduction on upstream information as

$$H(X_j|X_i)=r \text{ bits, } j \neq i, \text{ and } 0 \leq r \leq R.$$

Therefore, for each edge(i,j), the data correlation coefficient $cost_{ji\_correlation} = H(X_j|X_i) / H(X_j)$,

where $0 \leq cost_{ji\_correlation} \leq 1$.

When $cost_{ji\_correlation}=0$, data are strongly correlated; when $cost_{ji\_correlation} = 1$, data are independent. The data correlation model for sensor networks as given in [6,7,8,9].

D. Ant-Aggregation Algorithm

The Sensor network modeled as weighted graph G(V,E) where multiple source $src \in V$ and a single destination $d \in V$. The cost of edge indicates nodes within direct communication range and cost is an estimate measured as Euclidean distance. The input to algorithm is list of source nodes $src \in V$. The algorithm assigns ants to source nodes. The ants search the routes and communicate with other through pheromones. The node potential is an estimate of distance to destination. Each ant iterates to construct aggregation tree where internal nodes are aggregate points.

The ants either try to find shortest route to destination and terminates or finds closest aggregation point of the route searched by previous ants and terminates. The algorithm converges to local best aggregation tree. In order to find the global optimal aggregation nodes, the algorithm iterates on different permutations of source nodes with cost associated with each to converge in minimum cost. The algorithm converges to minimal cost aggregation points of aggregation tree. In order to converge, with early aggregation, weight is given to search the closest aggregation point. To search shortest path to destination weight is given to the distance estimates to destination. The node potential remembers estimate closest aggregation point or shortest route to destination. The traced edges are weighted using pheromone which communicates with other ants. The optimal path edges converge with minimal weights of pheromone. The non-optimal edges gets evaporated their pheromone. Thus, algorithm is governed by pheromone trails and node potential. The algorithm runs in two passes: Ants from source moves forward to destination searching the route to destination and ants takes reverse pass from destination to source updating pheromone trails over the edges and updating node potential as estimate for reaching closest aggregation or destination.

The governing equations for passes are given below:-

An artificial ant placed randomly in nodes and during each iteration chooses next node is governed by following formula:

Each ant located at node i hops to node j selected among the neighbors that have not yet been visited according to probability. Probability that ant k in node i will go to node j

$$P_k(i,j)=\left\{ \frac{\tau_{ij}^{\alpha} / d_{ij}^{\beta}}{\sum_{g \in J(i)_k} \tau_{ig}^{\alpha} / d_{ig}^{\beta}} \right\} \qquad (7)$$

α : relative importance of pheromone trial

β : relative importance of the distance

g : neighborhood of current node i

d_{ij} : node potential, gives estimate of early aggregation or shortest route to destination.

In the reverse pass each ant updates pheromone trails and node potential. The governing equations pheromone trials are follows:

$$\tau_{ij} = (1-\rho)\,\tau_{ij} + \rho * Q/totalcost \qquad (8)$$

where ρ and Q are ACO parameters and totalcost is obtained from first pass.

For other nodes ie not visited in first pass pheromone evaporates more rapidly for lower values:

$$\tau_{ij} = (1-\rho)\,\tau_{ij} \qquad (9)$$

The node potential is weighted function of either reaching nearest aggregation point or shortest route to destination or having high correlation updated as follows:

$$d_{ij} = \gamma * cost_{reaching\_aggregation} + \eta * cost_{destination} + \kappa * cost_{ij\_correlation} \qquad (10)$$

where γ and η are weights of choosing early aggregation or choosing route to destination ie distance estimates and κ is the weight for choosing data correlation.

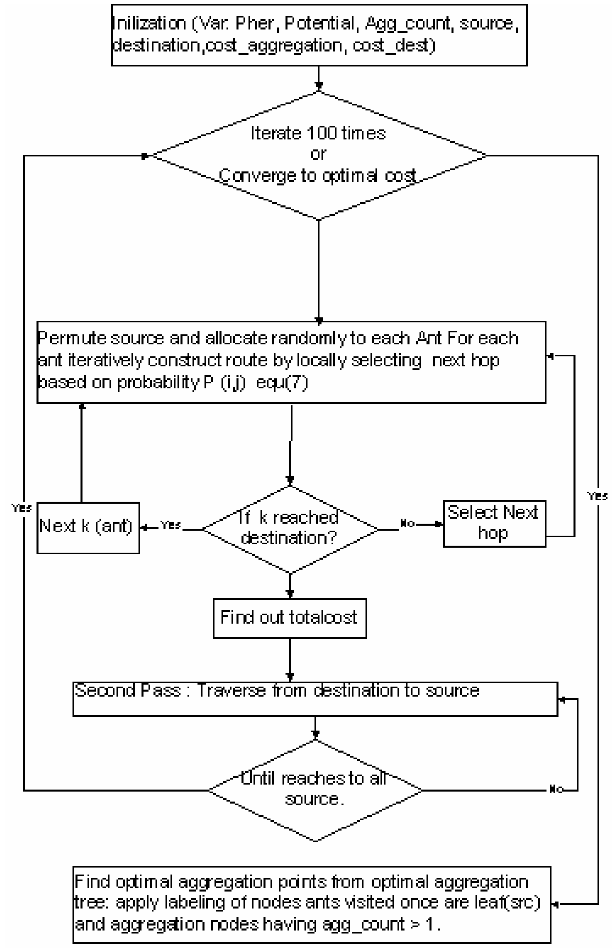

Fig.2: Ant-Aggregation algorithm

VII. IMPLEMENTATION & RESULTS

The Ant-Aggregation algorithm is simulated in MATLAB with a setting of sensor network of 50 nodes. The neighborhood is obtained from the random topology. Set of source nodes 20, 30&40 and a destination is considered for generating optimal aggregation tree using Ant-Aggregation. On varying the ACO parameters and weights of aggregation, shortest distance and correlation, the optimal aggregation tree in sensor network is obtained. The fig.3 below shows the convergence of optimal solution globally in the setup after 1000 iterations.

The ACO parameters in the simulation setup are given in table below:

ACO Parameters	Simulation Values
Q	8
ρ	0.8
α	3
β	6
γ	0.5
η	0.4
κ	0.1

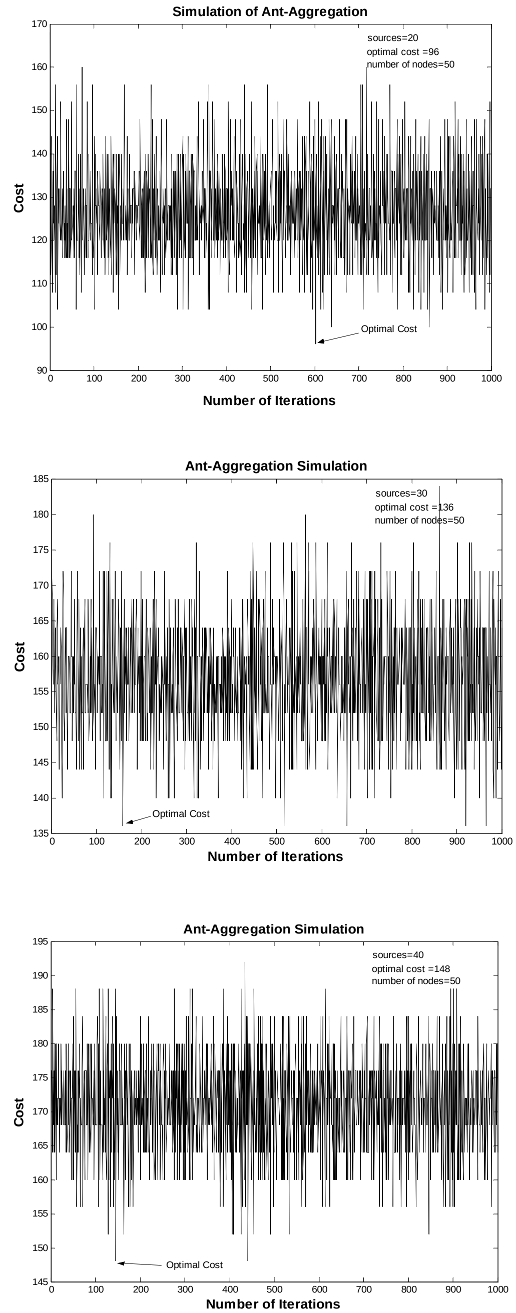

Fig. 3. Simulation of Ant-aggregation algorithm, convergence to optimal cost aggregation tree is pointed.

# of Source nodes	Opportunistic aggregation cost	Greedy aggregation cost	Optimal Aggregation cost
20(moderate # of source)	140(45%)	125(30%)	96
30	164(20%)	157(15%)	136
40(large # of source)	180(20%)	173(15%)	148

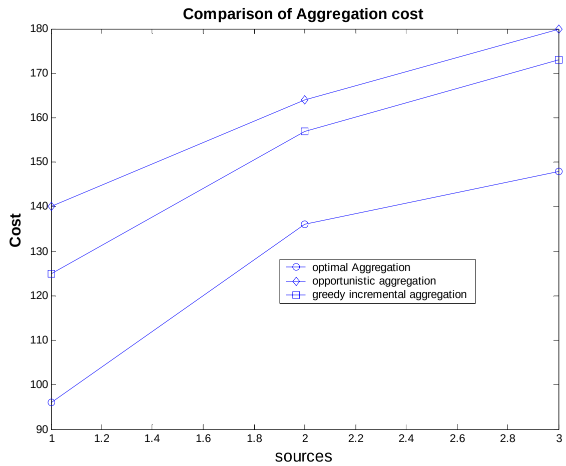

Fig.4 Comparison of aggregation cost

Energy-efficiency is measure as ratio of differential cost to the optimal aggregation cost.

Thus optimal aggregation saves energy 30%(on moderate number of source) to 15%(on large number of source) in greedy aggregation . Similarly optimal aggregation saves 45%(on moderate number of sources) to 20%(on large number of sources) in opportunistic aggregation..

It is shown that with less number of source nodes in network the energy savings in optimal aggregation is 45%-30% in comparison to other approximation schemes.

VIII. CONCLUSION

An ant based solution for the Optimal Aggregation Problem has been implemented and investigated. Extensive simulation is carried out for correctness of algorithm. It is observed that aggregation energy efficiency depends on the number of sources. The results of simulation reveal that optimal aggregation save energy upto 45% for moderate number of source nodes.

REFERENCES

[1] B. Krishanamachari, D. Estrin and S. Wicker, "*The Impact of Data Aggregation in Wireless Sensor Networks*", International Workshop of Distributed Event Based Systems (DEBS), Vienna, Austria, July 2002.

[2] C. Intanagonwiwat, D. Estrin, R. Govindan, and J. Heidemann. *"Impact of Network Density on Data Aggregation in Wireless Sensor Networks*", Technical Report 01-750, University of Southern California, November, 2001

[3] Utz Roedig, Andre Barroso, and Cormac J. Sreenan. "*Determination of Aggregation Points in Wireless Sensor Networks"*, Proceedings of the 30th Euromicro Conference (EUROMICRO2004), Rennes, France, pp. 503–510, IEEE Computer Society Press, August 2004

[4] S. Das, G. Singh, S. Gosavi, S. Pujar, "Ant Colony Algorithms for Data-Centric Routing in Sensor Networks", *Proceedings, Joint Conference on Information Sciences*, Durham, North Carolina, pp. 1565-1568, 2003.

[5] Al-Karaki, J. N., R. Ul-Mustafa and A. E. Kamal, "Data Aggregation in Wireless Sensor Networks - Exact and Approximate Algorithms", in the proceedings the International Workshop on High-Performance Switching and Routing, Phoenix, AZ, April 2004.

[6] R.Cristescu,B.Beferull-Lozano, and M.Vetterli, " On network correlated data gathering", Infocom'04,Hong Kong,2004.

[7] H.Cunqing, and Y.P.Tak-Shing, "Optimal Routing for Maximizing Lifetime of Wireless Sensor Networks",Infocom'05,2005

[8] D.Slepian and J.Wolf, "Noiseless coding of correlated information sources", IEEE Trans. Information theory,noIT-19,pp. 471-480,1973.

[9] S.Pradhan and K. Ramchandran, "Distributed source coding using syndromes(discus):design and construction", Proc IEEE Data Compression Conf.,March 1999.