Brief Announcement: Universal Data Aggregation Trees for Sensor Networks in Low Doubling Metrics*

Srivathsan Srinivasagopalan, Costas Busch, and S. Sitharama Iyengar

Computer Science Department, Louisiana State University, Baton Rouge, USA
{ssrini1,busch,iyengar}@csc.lsu.edu

Abstract. We describe a novel approach for constructing a single spanning tree for data aggregation towards a sink node. The tree is *universal* in the sense that it is static and independent of the number of data sources and fusion-costs at intermediate nodes. The tree construction is in polynomial time, and for *low doubling dimension* topologies it guarantees a $O(\log^2 n)$-approximation of the optimal aggregation cost. With constant fusion-cost functions our aggregation tree gives a $O(\log n)$-approximation for every Steiner tree to the sink.

1 Summary

We consider the fundamental problem of data aggregation in sensor networks towards a sink node s. The motivation comes from the hierarchical matching algorithm in [1] that constructs a data aggregation tree that is *simultaneously* good for all canonical fusion-cost functions f which are concave and non-decreasing. However, every time the data sources change the aggregation tree has to be reconstructed. In order to alleviate this problem, a novel approach in [2] builds a *single* aggregation tree which is independent of the number of data sources. This is for networks formed by randomly distributed nodes over a 2-dimensional plane. Here, we combine the above ideas to build a *universal* aggregation spanning tree for low (constant) doubling-dimension graphs, such that the tree is independent of the data sources, and can accommodate any canonical fusion-cost function. Our approach gives a $O(\log^2 n)$-approximation to the optimal aggregation cost (measured as the total involved communication cost), where n is the number of nodes. For constant fusion-cost we obtain a spanning tree that approximates within a $O(\log n)$ factor every Steiner tree that includes the sink.

Given a graph G, we first construct a *virtual tree* T based on a hierarchical clustering. The performance of T is estimated for low doubling-dimension graphs. We then project T to a real spanning tree that preserves the asymptotic aggregation performance. We provide a hierarchical clustering $Z_0, \dots, Z_\kappa$ with $\kappa + 1 = O(\log n)$ levels, where each Z_i is a partition of G. Every cluster in Z_i is a connected subgraph of G with diameter $O(2^i)$. Each cluster has a leader; let I_i

* Supported by PKSFI grant 32-4233-40860-LSU, and NSF grant 0846081.

S. Dolev (Ed.): ALGOSENSORS 2009, LNCS 5804, pp. 151–152, 2009.

denote the leaders at level i. For any $u, v \in I_i$, $\text{dist}_G(u, v) = \Omega(2^i)$. Each cluster at level i (with leader ℓ_i) is completely contained by some cluster at level $i+1$ (with leader ℓ_{i+1}), with the exception of $i = \kappa$. Virtual tree T consists of $\kappa + 1$ levels and has as nodes the leaders of the clusters, edges of the form $e = (\ell_i, \ell_{i+1})$ (ℓ_i is the child), and weight of e proportional to $\text{dist}_G(\ell_i, \ell_{i+1})$. The root of T is s (which is the leader of the single cluster of Z_κ), while the leaves are the individual nodes of G (partition Z_0 consists of the individual nodes of G).

The data aggregation in tree T is performed level by level in κ phases. In phase i each leader at level i receives data from its children and then fuses the data. The size of the data produced by a node during aggregation depends on the fusion-cost function f. Assume for now a constant function $f(\cdot) = 1$. For the analysis consider an arbitrary set of source nodes A. Let $X_i \subseteq I_i$ be the set of leaders at level i which hold data at the end of phase i. In phase $i + 1$, each leader in X_i sends a message to its parent giving a total communication cost $O(|X_i| \cdot 2^{i+1})$. For every leader $v \in X_i$, there is at least one node in A that belongs to the cluster of v at level i. Let $A_i \subseteq A$ denote the union of such data sources, where $|A_i| = |X_i|$. Using the doubling dimension property of G it can be shown that there is a subset $A'_i \subseteq A_i$ such that $|A'_i| \geq |A_i|/(2^{3\rho} + 1)$ (where ρ is the doubling dimension parameter), and for any pair $u, v \in A'_i$, $\text{dist}_G(u, v) = \Omega(2^i)$. The optimal aggregation cost for A is at least the cost of the Steiner tree connecting the A'_i to the sink which is $\Omega(|A'_i| \cdot 2^i) = \Omega(|A_i| \cdot 2^i) = \Omega(|X_i| \cdot 2^i)$ (for constant ρ). Therefore, the aggregation cost of a phase is optimal within a constant factor. Considering all $\kappa = O(\log n)$ phases, the total aggregation cost is within a factor of $O(\log n)$ from optimal.

For the analysis of non-constant fusion-cost functions, we further divide set X_i into $\lambda = O(\log n)$ classes, $X_i^0, \ldots, X_i^{\lambda-1}$, so that at the end of phase i each node $v \in X_i^j$ contains data of size $f(|A_i^v|) \in [2^j, 2^{j+1})$, where $A_i^v \subseteq A$ is the set of source nodes which belong to the cluster of v at level i. By collapsing all the nodes in A_i^v to a single artificial data source of weight 2^j, the data aggregation problem of each class is reduced to the constant fusion-cost function case (an additional $\Theta(2^j)$ factor appears in both the lower and upper bounds). Thus, the aggregation cost for each class is optimal within a constant factor. By considering all classes and phases we obtain a $\kappa\lambda = O(\log^2 n)$ approximation.

The projection of the virtual tree to a real spanning tree is performed by replacing every edge in T with a shortest path in G. The aggregation performance is maintained by avoiding interference between paths of the same level.

References

[1] Goel, A., Estrin, D.: Simultaneous optimization for concave costs: single sink aggregation or single source buy-at-bulk. In: Proceedings of the fourteenth annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2003), Baltimore, Maryland, pp. 499–505 (2003)

[2] Jia, L., Noubir, G., Rajaraman, R., Sundaram, R.: Gist: Group-independent spanning tree for data aggregation in dense sensor networks. In: Gibbons, P.B., Abdelzaher, T., Aspnes, J., Rao, R. (eds.) DCOSS 2006. LNCS, vol. 4026, pp. 282–304. Springer, Heidelberg (2006)