Voronoi Trees for Hierarchical In-Network Data and Space Abstractions in Wireless Sensor Networks

Mohamed M.Ali Mohamed
University of Illinois at Chicago
Chicago, IL. USA
mali25@uic.edu

Ashfaq Khokhar
University of Illinois at Chicago
Chicago, IL. USA
ashfaq@uic.edu

Goce Trajcevski
Northwestern University
Evanston, IL. USA
g-trajcevski@nothwestern.edu

ABSTRACT

We address the problem of spatial queries in Wireless Sensor Networks (WSN) via hybrid overlays, where the data values may correspond to different physical phenomena and sensors may be correlated with spatial constraints. We show how hierarchical data and space abstractions can be used to represent Voronoi Cell based partitions of the sensing field and use Voronoi Trees to efficiently map the hierarchical abstractions for energy-efficient processing. The proposed scheme is simulated on the SidNET, a JiST/SWANS based sensor network simulation platform. The performance results show significant advantages in terms of accurate field representation at different levels of the tree hierarchy with a trade-off in query processing delay.

Categories and Subject Descriptors

C.2.4 [**Distributed Systems**]: Distributed Applications; E1 [**Data Structures**]: distributed data structures, graphs and networks.

General Terms

Algorithms, Design, Experimentation.

Keywords

Wireless Sensor Networks, Computational Geometry.

1. INTRODUCTION

Wireless Sensor Networks (WSN)s have become tools for distributed data gathering and query processing [1] in a plethora of application domains. Retrieving a perfectly accurate (spatial distribution of) data is expensive in terms of energy usage and processing time, and is not always needed [1]. We investigate issues related to distributed in-network indexing and query processing, taking into consideration that due to the imperfections of sensing devices, approximate responses with bounded error may be acceptable in various contexts. Since reduction in accuracy can also decrease resources utilization, we aim at designing abstractions for answering queries pertaining to overlapping heterogeneous data and/or spatial regions, at various levels of accuracy.

Traditional indexing structures (K-D trees, R-Trees, B-Trees, or variations) used for efficient retrieval in spatial databases [3] have lead to distributed index structures for WSNs [5-9]. However, the particular challenge that we address in this work is efficient in-network processing of hybrid queries, by minimizing communication, computation, and storage overheads. These queries combine sensed values within a specific geographic region, and the spatial extent at which the sensed data falls within a specific range, like, for example:

Q1: *Retrieve all the prairie regions in DuPage county that are near river and have between 15% and 45% of salinity decline in the next 3 hours.*

Motivated by the well-studied properties of Voronoi diagrams from Computational Geometry [4] and the fact that a Voronoi cell centroid is optimal for sensor allocation in terms of sensing accuracy [12], in this paper we investigate the use of Voronoi Treemaps [2] as a hierarchical indexing structure in WSN.

Voronoi tessellations have been used for tackling various objectives in WSN research, where the data-centric nature of the underlying sensed phenomena matches some basic properties satisfied in a region bounded by a Voronoi cell [12]. Voronoi Treemaps create hierarchical spatial data structure that splits the given space in multiple granulation levels by creating contiguous non-overlapping regions, and without producing holes.

Our proposed techniques are carefully designed to accommodate different query types and provide responses within specified accuracy, incurring diminutive update costs that do not increase across the network. This enables load balancing and prolonging the overall network lifetime. We use rank order statistics, regular sampling, and bitmap representations to interleave the management of physical- as well as data space queries. The proposed techniques are implemented on SIDnet simulator [11]. Our experimental observations show the benefits and trade-offs.

2. PRELIMINARIES

We assume a dense WSN where each node s_i knows its coordinates (x_j, y_j) with respect to a reference coordinate system, and the locations of all of its 1-hop neighbors. Users may be interested in the sensed phenomena across the field or parts of it, but the location of sensed values is also important as the value itself – , e.g., fire occurrence requires specifying where this event is detected. In this work, we are focusing on two main types of queries: (1) inquiring the values of sensed phenomenon (physical-space queries), and (2) inquiring the location of sensor nodes with certain sensed values (data-space queries). Queries may be confined to a regional space of the field or a range of values – a constraint term specifying the search bound. Allowing limited (im)precision of the answer to decrease resources utilization, we focus on conjunction of query predicates, pertaining to heterogeneous types and (sensed) values.

MSWiM '13, November 03 - 08 2013, Barcelona, Spain
Copyright 2013 ACM 978-1-4503-2353-6/13/11…$15.00.
http://dx.doi.org/10.1145/2507924.2507999

2.1 Voronoi Diagrams and Hierarchies

One of the most studied concepts in Computational Geometry [4] is the Voronoi diagram of a collection of points, along with its geometric-dual, Delaunay triangulation. Given n distinct points $P = \{p_1, p_2, \ldots, p_n\}$ contained inside a convex polygon S in R^2, the bounded Voronoi tessellation $V_{\cap S}(P) = \{V(p_1)\cap S, V(p_2)\cap S, \ldots, V(p_n)\cap S\}$ is defined as the decomposition of the space S into n convex polygons, called Voronoi regions. Each Voronoi region $V(p_i)$ associated with the point p_i, has the property that for an arbitrary point $q(x,y)$, if $q(x,y)$ is inside $V(p_i)$, than $dist(q, p_i) > dist(q, p_j)$, for any other $p_j \in P$ $(i \neq j)$. The function *dist* represents a specified distance measure between two points, which can be as simple as Euclidean distance, or may include other functions that parameterize more context-related information. A specific example is the centroidal Voronoi tessellation (CVT) - a Voronoi tessellation with the property that the generating point for each Voronoi region is the center of mass of the region.

A bottom-up approach for constructing a hierarchy based on Voronoi cells was presented in [13], where at each level of the hierarchy, a Voronoi tessellation is calculated such that each set of Voronoi cells in a given level is contained in one Voronoi cell in the higher level. However, the boundaries of the Voronoi cells between levels are not guaranteed to fully coincide, resulting in non-containment of lower level cells in higher levels.

Tree-Maps [10] visualized hierarchical information structures by mapping the full hierarchy onto a rectangular region in a space-filling manner. Voronoi Treemaps [2] build a hierarchical spatial data structure that is based on CVTs in a top-down manner, partitioning a given convex polygon space into a set of centroidal Voronoi cells, and recursively partitioning each cell into further CVT, given the generating points.

A given space is split in multiple granulation levels by creating contiguous non-overlapping regions without producing holes.

3. ABSTRACTIONS AND MODELS

We firstly present data and physical space abstractions that will be used to represent sensor data and its location at different levels of the Voronoi Treemap (see [5] for a preliminary version), in the context of orthogonal bisection based KD-trees. We then follow with a detailed discussion on Voronoi-based partitioning.

3.1 Physical-space Abstraction

Our focus is on representing the sensed data in the field of interest at multiple scales with respect to the geographical location of the sensing nodes. Given a hierarchical spanning tree $T(V',E')$ of depth d, and *(w.l.o.g.)* a fixed fan-out f, in a graph $G(V, E)$, where each leaf node $v_{d,i}$ represents a Voronoi Cell, we want to find an abstract representation $p_{v'di}$ of the sensed values in the region associated with node $v'_{d,i}$, such that: *average error*, *communication cost*, and *computation cost* are minimized. A more formal definition can be found in [5].

Each node $v'_{d,i}$ in the spanning tree $T(V',E')$ gathers the sensed from its set of sensor nodes v_{ij} in its vicinity and stores them. Upon acquiring its population's readings, each node $v'_{d,i}$ rank-orders the sensed data and stores it in an array along with their corresponding sensor nodes' physical locations. The abstract representation $p_{v'di}$ is introduced as an array of a fixed-size (k) and its values are chosen by regularly sampling the sorted array of the population nodes with interval (D/k), including the first and last elements of the sorted array, as depicted in Figure 1.

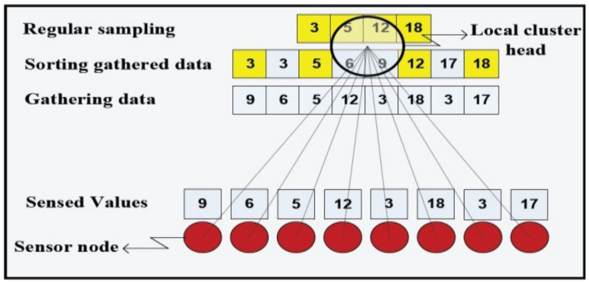


Figure 1. Leaf (cluster-head) create the representation construct by sorting and sampling sensed values.

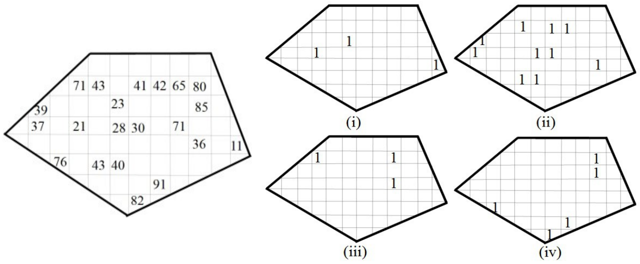
To develop abstract representations of larger regions represented by the non-leaf nodes of the spanning tree, each leaf node $v'_{d,i}$ compresses and transmits the representation $p_{v'di}$ to its parent in the spanning tree. Each non-leaf node receives from its children a set of f arrays, representing the physical-phenomenon at spatially non-overlapping regions in the field, and merges them into a larger one representing the sensed phenomenon in the area enclosing the f regions. The new physical-space abstraction of the larger region is created by regularly sampling the merged array with a given sampling interval (f). Upon completion, the new construct is of the same size as the received (input) arrays. It provides a regular sorted sample of the larger population, but in a coarser form. The data-payload is updated throughout the participating nodes by using fixed size messages, keeping the communication workload equally distributed.

3.2 Data-space Abstraction

We assume that the possible values of the sensed phenomenon are delimited within a finite range [min, max] for the potentially queried data-space. Our objective is to provide a hierarchical multi-resolution abstraction scheme to obtain approximate answers to queries that involve localizations of the related sensor nodes. The data-space abstraction aims to find a representation construction $L_{v'di}$ of the sensor nodes location distribution *w.r.t* its sensed value for each leaf node $v'_{d,i}$, such that the *average error*, *communication cost*, and *computation cost* are decreased.

Each set of sibling leaf nodes represents a group of neighboring clusters within a region splitting the data-space into q data ranges, where $q \geq f$. The responsibility of the data-space in each region gets distributed among the managing f leaf nodes. Each cluster head node (i.e., leaf node $v'_{d,i}$) keeps the position of the sensors which report value within its data range(s) of responsibility inside its containing region. For example, for a sensed phenomenon whose possible range of values is [1-100], a split between four nodes is: {[1,25], [26,50], [51,75], [76,100]}. Thus, each leaf node reports to its f-1 siblings the positions of the nodes of its population conforming to their assigned data range.

In order to create the representation construct $L_{v'di}$ of nodes positions for each leaf node $v'_{d,i}$, a bit-map is created. A bit-map is a is a 2D array of a size that maps to the physical region it represents, where each entry represents an area that can be occupied by no more than a single sensor node. We note that the resolution of the map (i.e. cell size) is application dependent. The map entries/cells are initialized to zero. A leaf node, then, sets the cells occupied by sensor nodes that reported values within its data range. The resulting map can be viewed as a highly sparse 2D array of zeroes and ones, which is compressed using Run-length coding technique. Figure 2 depicts an example of bit-map construction for the sensed values within one region.



(a). Sensed values in a range [1,100] reported in a region.

(b). Corresponding four bitmaps for ranges [1,25], [26,50] [51,75] and [76, 100], respectively.

Figure 2. Illustrative example of the bitmap creation.

For each data range, the corresponding responsible nodes (*f* in total) at level *i* send their compressed maps to the node in level *j* responsible for the same data range in the containing region. Upon receiving the *f* maps for the data range, the recipient node concatenates them according to their geographic locations, which generates one larger map for the data range of responsibility in the whole region.

In order to provide approximate representation and keep message size fixed across the indexing structure, the concatenated map needs to be zoomed out with a scale that reduces its size with a *1/f* factor on average. In this fashion, nodes keep approximating maps of different data ranges as they elevate through the indexing structure, providing the ability to supply proper approximations for the data-space as shown in Figure 3.

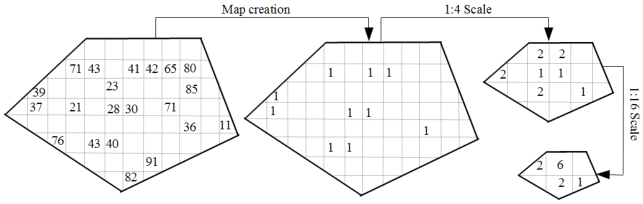


Figure 3. Map creation for data range [26-50], and zooming out twice with 1:4 factors.

3.3 Voronoi Treemaps and Queries

The Voronoi treemap (cf. Sec. 2.1.) is created using a top-down approach, in which the field is subdivided into a set of centroidal Voronoi cells, according to a given set of generating points which are selected via a given function having prior knowledge of the distribution of the sensed phenomenon. Within each cell, a node is placed at the centroid point and considered responsible of managing the data abstraction processes, and reporting them to the sink node. The same process is repeated within each centroidal Voronoi cell, to create the next level of spatial division in the field corresponding to the next level of indexing nodes in the Voronoi Treemap. This recursion stops at the last level of Voronoi cells – local clusters - surrounding the phenomenon sources. After constructing Voronoi treemap, sensor nodes are deployed in each local cluster surrounding the cell centroid point, which has the local cluster head node, and represents the source of the sensed phenomena in the cell. Figure 4 shows a color-coded description of the elected nodes at 3 inner-levels of a Voronoi Treemap. Leaf nodes create maps for the data ranges of their regions, transmit them to the upper level nodes responsible for the same data range. Upper level nodes zoom-out the maps and recursively apply the same process until the sink node is reached. Red star for the data source producing higher sensed values than yellow star**.** Queries originate at a sink, which we assume is connected to a base station. Upon receiving a query, the root node first checks the query type to decide which representation is inquired. It then analyzes the bounding constraints –if any exists– for range queries, where through the bounding region and bounding data range constraints the node can determine which nodes need to participate in the processing. The solution path is determined step by step, where each node that receives the query checks the intersection of the query's geometric and data range bounds with the geometric area and data range(s) that it covers. The decision of being able to answer the query at each node is taken according to the accuracy requirement of the query.

Once the query reaches the node(s) capable of answering it with satisfactory accuracy, the response is backtracked through the same path it took from the root node. Each intermediate node waits to receive the response from all the nodes it forwarded the query to. Once received, it combines data-space query responses. The query results are obtained by concatenation of the response maps, while physical-space range queries merge the arrays of sensed values. In the case of extreme values (min or max) the merging trivially preserves the smallest or largest values.

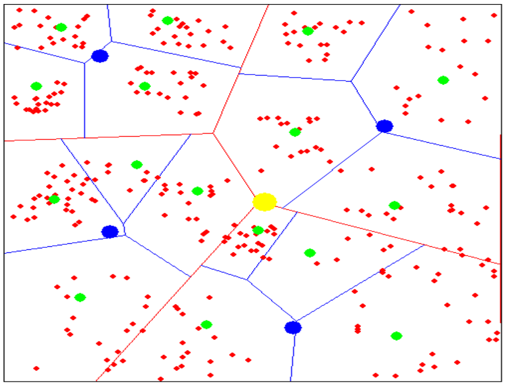


Figure 4. Color-coded description of the elected nodes at different levels of. Yellow (largest) node is the sink. The four blue nodes are the next level, and the 16 green nodes are local cluster-heads. Sensor nodes are the tiny red circles.

4. EXPERIMENTAL OBSERVATIONS

We implemented the Voronoi Trees and protocol on top of SIDnet-SWANS [11] WSN simulator (based on Jist-SWANS event simulation engine: http://jist.ece.cornell.edu/index.html). An orthogonal bisection based KD tree implementation in SIDnet (cf. [3]) was used to compare the performance of Voronoi treemaps. While there are many different binary space partitioning methods, we opted to use KD-tree because, to the best of our knowledge, there are no indexing methodologies in the literature that encompass data space abstractions. Hence, we resorted to our own implementation for this, as well as hybrid-settings. The WSN has 500 nodes deployed in a square field of 300x300 meters square, using MAC802.15.4, and Shortest Geographic Path for routing. The power consumption characteristics are based on Mica2 Motes specifications, MPR500CA.

Phenomenon Values Impact: The quality of representing the sensed values via hierarchical abstraction(s) is limited by the ability of the sensors to report values that match the distribution of the sensed phenomenon. For this, sensor nodes should be located in positions that correspond to the sensed phenomenon distribution across the field. We tested the indexing structures in terms of accurately representing data field. We simulated the data field using 2D Gaussian distributions by varying their variance. We reconstruct the field by looking at the variance at different levels of the hierarchy in these indexing structures. Figure 5 shows that the Voronoi Treemaps achieve lower error when reconstructing "sharp" Gaussian distributions (i.e., lower variance) than KD trees. This is intuitive in the sense that the Voronoi cells are created based on the properties of the physical field being sensed. Hence, closer mapping of sensor nodes to data sources yields reconstructed values that are closer to the phenomenon's spatial distribution.

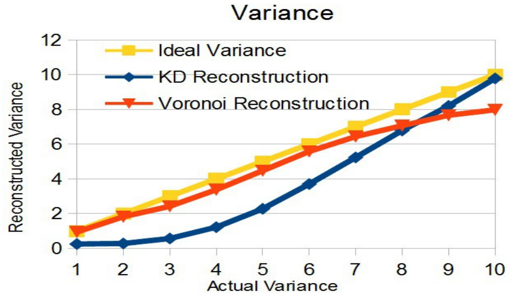


Figure 5. Reconstructed Variance

Similar observations are present in Figure 6 which shows the normalized error values. On the contrary, random deployment on which the KD tree is built is more likely to have less accuracy in the reconstructed data field. This will be particularly true for sharp Gaussian distributions. For distributions with significantly larger variances, the KD tree starts to perform better in reconstructing the field than the Voronoi treemap. The reason is that Voronoi treemap deploys nodes around the data source, whereas KD-trees, covering larger area between children and cluster head, may incorporate the readings from sensors that are away from the source.

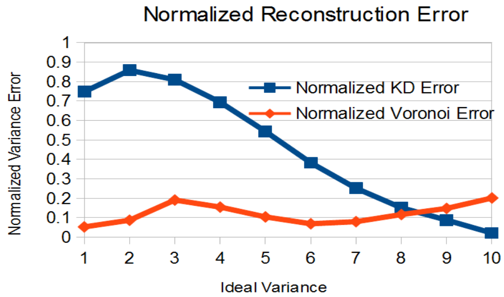


Figure 6. Normalized Variance Reconstruction Error

5. CONCLUSION AND FUTURE WORK

We presented a hierarchical data indexing and querying system based on Voronoi Treemaps, using novel abstraction techniques for data and physical space based on rank order sampling and bitmaps representation. The abstraction schemes are applicable to a wide range of applications because of their generic nature. The data-space abstraction based on bit maps is first of its kind in indexing in wireless sensor networks. Voronoi Treemaps provide sensor nodes deployment that is better capable of capturing the phenomenon distribution. The proposed system gives a generic way to answer different queries of various phenomena with high efficiency. Fixed size update messages ensure load balancing across the network, enabling longer network life time. Our ongoing efforts are focused on incorporating spatio-temporal evolution of the Voronoi treemaps for WSN with mobile sensors.

6. REFERENCES

[1] Zhao, F. and Guibas, L. 2004. *Wireless Sensor Networks: An Information Processing Approach*. Morgan Kaufman.

[2] Balzer, M., Deussen, O. and Lewerentz, C. 2005. Voronoi treemaps for the visualization of software metrics. *In Proc. of* ACM Symp. on Software visualization.

[3] Shekhar, S. and Chawla, S. 2003. *Spatial Databases: A Tour*. Prentice Hall.

[4] De Berg, M., Cheong, O., Van Kreveld, M. and Overmars, M. 2008. *Computational Geometry Algorithms and Applications*. Springer.

[5] Ali Mohamed, M. M., Khokhar, A., Trajcevski, G., Ansari, R. and Ouksel, A. 2012. Approximate hybrid query processing in wireless sensor networks. *In Proc. of* ACM GIS.

[6] Greenstein, B., Estrin, D., Govindan, R., Ratnasamy, S. and Shenker, S. 2003. DIFS: A Distributed Index for Features in Sensor Networks. Ad Hoc Networks, Vol. 1, pp. 333-349.

[7] Ganesan, D., Greenstein, B., Estrin, D., Heidemann, J. and Govindan, R. 2005. Multiresolution Storage and Search in Sensor Networks. ACM Transactions on Storage (TOS), Vol. 1, pp. 277-315.

[8] Ciancio, A., Pattem, S., Ortega, A. and Krishnamachari, B. 2006. Energy-Efficient Data Repres and Routing for Wireless Sensor Networks Based on a Distributed Wavelet Compression Algorithm. IPSN.

[9] Meliou, A., Guestrin, C. and Hellerstein, J. 2009. Approximating Sensor Network Queries Using In-Network Summaries. IPSN Johnson, B. and Shneiderman, B. 1891.

[10] Tree-maps: A space-filling approach to the visualization of hierarchical information structures. In Proc. IEEE Visualization Conf.

[11] Ghica, O., Trajcevski, G., Scheuermann, P., Bischoff, Z. and Valtchanov, N. 2008. Sidnet-swans: A simulator and integrated development platform for sensor networks applications. ACM SenSys.

[12] Ammari, H.M., and Sajal K.D. 2008. Promoting heterogeneity, mobility, and energy-aware voronoi diagram in wireless sensor networks. IEEE Trans. Parallel and Distributed Systems, IEEE Transactions on 18(7)

[13] Gold, C. and Angel, P. 2006. Voronoi Hierarchies. Geographic Information Science (pp. 99-111). Springer Berlin Heidelberg.