# Coping with irregular spatio-temporal sampling in sensor networks

Deepak Ganesan[†] Sylvia Ratnasamy[†‡] Hanbiao Wang[†] Deborah Estrin[†]

† Department of Computer Science, UCLA, Los Angeles, CA 90095

†‡ Intel Research, Berkeley, CA 94704

{deepak@cs.ucla.edu, sylvia@intel-research.net, hbwang@cs.ucla.edu, destrin@cs.ucla.edu}

## Abstract

Wireless sensor networks have attracted attention from a diverse set of researchers, due to the unique combination of distributed, resource and data processing constraints. However, until now, the lack of real sensor network deployments have resulted in ad-hoc assumptions on a wide range of issues including topology characteristics and data distribution. As deployments of sensor networks become more widespread [1, 2], many of these assumptions need to be revisited.

This paper deals with the fundamental issue of spatio-temporal irregularity in sensor networks We make the case for the existence of such irregular spatio-temporal sampling, and show that it impacts many performance issues in sensor networks. For instance, data aggregation schemes provide inaccurate results, compression efficiency is dramatically reduced, data storage skews storage load among nodes and incurs significantly greater routing overhead. To mitigate the impact of irregularity, we outline a spectrum of solutions. For data aggregation and compression, we propose the use of spatial interpolation of data (first suggested by Ganeriwal et al in [3]) and temporal signal segmentation followed by alignment. To reduce the cost of data-centric storage and routing, we propose the use of virtualization, and boundary detection.

## 1 Motivation

Wireless Sensor Networks have received tremendous attention over past few years. Early research in this area [4, 5, 6] has identified several important research challenges: energy efficiency, system and environmental dynamics, resource constraints, calibration, etc . These fundamental challenges have led to exciting research in data aggregation ([5, 7]), self-configuration ([8, 9]), distributed storage ([10]), GPS-less localization and time-synchronization [11], among others.

A central issue that cuts across each one of these research thrusts is the impact of irregular sampling. We argue that spatio-temporal irregularity is fundamental to wireless sensor networks and must be considered by sensor network algorithm, protocol, and application designers. We contend that this bit of reality is as fundamental to sensor network development, as was the adoption of bursty traffic models for the design and analysis of Internet protocols. The remainder of this introduction describes the origin of irregular sampling in wireless sensor networks.

**Spatially Irregular Deployments**

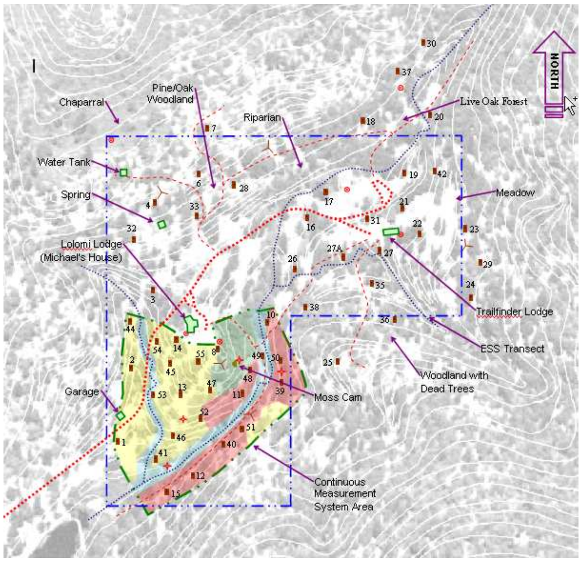
*Can we expect sensor network deployments to be uniformly regular?*

Most sensor network deployments will have irregular spatial configurations for two fundamental reasons: (a) the phenonmena of interest are not uniformly distributed and the deployment of sensor resources will be variable in order to achieve denser sensing where there is greater spatial variability (e.g., on the edge of biological regions), and (b) terrain and other deployment practicalities bias deployment locations to where necessary power sources, communication or access can be achieved.

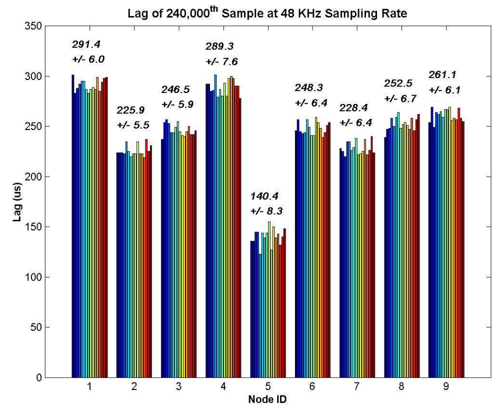
Sensor networks in built environments such as structures and factories might be deployable with regular topologies. However, in environmental monitoring networks such as that shown in Figure 1(a), node placement is highly irregular, both because of the ecological interest in particular biotic regions, and because of terrain conditions that render uniformly random topologies unachievable.

*Why are irregular topologies a concern?*

Irregular deployments impact many aspects of sensor network design and performance. For example, spatial sampling of data depends on the topology, hence, data processing schemes that take into account spatial location or frequency will be affected. Consider the common signal-processing problem of ensuring Nyquist sampling of a band-limited signal to prevent aliasing. Such a problem becomes more complex in the distributed sensor network case, since the system needs to ensure Nyquist sampling rate in the presense of network dynamics such as




(a) Micro-climate monitoring sensor network deployment at James Reserve: Node placement is irregular, with the lower left being more densely deployed than the rest of the network.




(b) This figure shows the lag between nine iPAQs that start sampling simultaneously and continue for 5 seconds at 48 KHz. The time difference of the last samples for different iPAQs can be up to 150 micro seconds.

Figure 1: Examples of spatio-temporal irregularity in sensor networks

node failures. Frequency-independent aggregation functions are impacted by irregular spatial sampling as well. Consider the simple problem of finding the average temperature in a spatial area. To provide an accurate result in an irregular setting, the spatial average needs to weigh nodes in a sparse area higher than nodes in a dense area (described in [3]).

Another architectural component that is impacted by spatial irregularity is data storage schemes that make use of uniform hashing over a name-space, followed by geographic routing using GPSR [12] to the node closest to the hashed location. Such schemes, that are loosely based on Data Centric Storage [10], move from a distributed storage model in which nodes store only locally sensed data to one in which this local data is stored at (possibly multiple) remote nodes. Hence the load (storage, lookup, routing etc) at a particular node is now impacted by the behavior of an arbitrary collection of remote nodes without knowledge of their network parameters (eg: storage capacity, bottleneck capacity, dynamism). If sensor networks were uniform in terms of placement, node capability, dynamics and sensed data, this lack of knowledge is not a concern, but irregularity can exacerbate the load and cost imbalance between different parts of the network.

**Temporally Irregular Sampling:**

*Can we expect clocks at different nodes in a sensor network to be continually, synchronized to the precision required by the application?*

Regular temporal sampling requires synchronized clocks at all of the measurement points. This is particularly an issue at intended deployment sites where GPS access is unavailable, for instance, much of the deployment shown in Figure 1(a) is in thick foliage. Similarly, seismologists have interest in studying wave propagation in canyons (such as San Gabriel) which do not have GPS access.

Recent research into time-synchronization for such GPS-less sensor networks([11, 13]) have shown that distributed, precise synchronization is indeed feasible. Such a procedure, however, comes with the associated cost of transmitting periodic beacons for noise reduction and multi-hop synchronization. Thus there is a fundamental tradeoff: more energy is required for finer synchronization for high-frequency sensing (e.g., seismic applications which sample at 100Hz, acoustic at 48KHz), while nodes with constrained energy budgets must emphasize energy conservation. For instance, Figure 1(b) shows that a cluster of ipaqs lose synchronization upto $150\mu s$ ($5$ acoustic samples) within a matter of seconds, requiring high rate of time-synchronization and therefore, high communication overhead. As pointed out by Elson et al. ([11]), variability in time-synchronization will result from precisely this need to sacrifice synchronization guarantees for energy consumption. Additionally, many external factors will contribute to such variability including changing ambient and system noise levels, loss of synchronization beacons, node failures and other system dynamics.

*Is imprecise and variable synchronization a concern?*

Variable time-synchronization impacts in-network data processing schemes such as collaborative signal processing (CSP) that rely on having precisely time-stamped samples for purposes such as target detection, localization, and data compression. For instance, consider an in-network processing scheme where a sensor node combines acoustic measurements at the nine ipaqs in Figure 1(b) to localize the sound source and point a camera accordingly. The synchronization error in this example (maximum of $150\mu s$) translates to localization errors upto $5$ cm for measurements within barely a few seconds of perfect synchronization of all Ipaqs. Further difference between measurement time and synchronization time will result in correspondingly greater error.

**Paper organization:** In the rest of this paper, we take an in-depth look at the impact of spatio-temporal irregularity on two building blocks of sensor network applications: data compression and data storage. For each of these subjects, we suggest techniques that can be used to tackle irregularity. We conclude with a discussion of a other sensor network research areas such as network connectivity that are impacted by irregularity.

# 2 Impact of sampling irregularity on data compression

One of the foremost research challenges in sensor networks is the disparity between the amount of data generated by sensors and the amount of data that a network can communicate before depleting limited energy resources or exceeding available link bandwidths in multihop systems. In-network data processing mechanisms seek to reduce the communication overhead of such networks by a wealth of different techniques such as exploiting correlation to compress data and performing data-interpretation ([14]) within the network.

Significant research has gone into distributed data aggregation. Various query types have been studied including standard SQL queries such as COUNT, MIN, MAX, AVG, SUM ([7, 3, 15]), more complex spatial features such as edges [14, 15], distributed compression and estimation techniques [16], etc. Most aggregation operators have to deal with irregularity in some phase of their processing. Consider, for instance, a COUNT aggregator, that determines the number of targets in a region. While such an aggregator seems unaffected by irregularity, the collaborative signal processing for target detection has to deal with temporal irregularity to detect or localize the target.

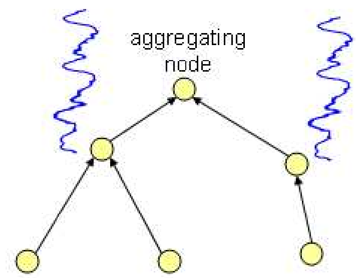



Figure 2: Section of an aggregation tree: Junction node compresses correlated spatio-temporal data from different areas of the network. Spatial irregularity requires interpolated coding schemes, and temporal irregularity (mis-aligned time-series) requires schemes to correct for time-varying lag.

For lack of space, we do not exhaustively deal with the impact of irregularity on all these data aggregation operators. Instead, we focus on one instance, data compression, whose performance is affected by both spatial and temporal irregularity. However, we believe that the techniques that we propose to deal with irregular data compression are also applicable to other aggregation operators that process distributed sensor data.

A typical application of data compression is in data gathering in sensor networks, where many sensors transmit data to a sink over a tree structure (a part of which is shown in Figure 2). In this example, we will assume that each junction node on this tree receives compressed data from its children, decodes the data, jointly compresses the decoded data with its own data, and forwards the encoded data to its parent on the tree. Our goal is not to target a specific coding scheme ([15, 17, 16] but to discuss practical concerns that are raised by irregular spatio-temporal sampling, and to suggest some initial solutions for these.

Spatial irregularity impacts an in-network compression schemes for data gathering simply because most practical codecs assume regularly sampled datasets. For instance, typical wavelet codecs such as JPEG operate on regularly sampled images whereas in-network compression using wavelets will need to handle spatially irregular data samples. Similarly, compressing two mis-aligned but correlated time-series typically reduces compression efficiency. For instance, in a simple test case, a bird call recorded at proximate Ipaq sensor nodes were compressed together using Delta coding followed by Huffman coding. In this case, a lag of 4 samples between the two time-series reduced compression efficiency by approximately 25%. This factor will be higher for more correlated datasets and greater lag.

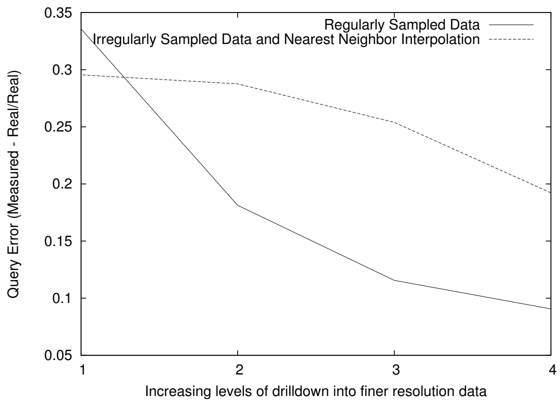



Figure 3: Performance of a drill-down MAX query. Irregular topology with nearest neighbor interpolation clearly performs significantly worse than one with regularly sampled data.

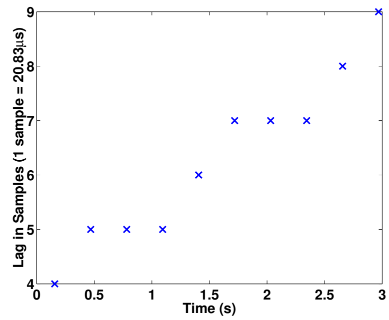



Figure 4: The lag between samples corresponding to a stationary acoustic source at proximate time-synchronized Ipaqs increases with time due to differences in clock frequencies between the two nodes. Cross-correlation techniques can correct for a constant lag but time-varying lag is significantly harder to deal with.

## 2.1 Coping with irregularity in data compression

Although the combination of irregular sensor data sampling and in-network processing is a novel challenge, irregularity has been dealt with extensively in contexts such as signal processing, geo-spatial data processing and computational geometry. In this section, we identify techniques in these diverse areas that can be applied and extended to a distributed setting to cope with irregularity.

**Interpolation of Spatially Irregular Data**

Irregular spatial samples are routinely regularized in geo-spatial data processing since analysis of irregular datasets is significantly more complex than that of regularly spaced ones. This regularization procedure, called *resampling*, typically involves interpolation and can be used to deal with irregularity.

While there are a wide range of interpolation schemes (polynomial, fourier, least squares, etc [18]), many of these schemes are not applicable for spatial interpolation in sensor networks due to their communication complexity. The cheapest interpolation scheme for distributed sensor data is *nearest neighbor* [18], which assigns the value of a resampled grid point to the nearest known data sample. Such sampling can be done in a distributed and inexpensive manner by constructing the Voronoi cells corresponding to each sensor node.

Such nearest neighbor (or k-nearest neighbor) interpolation techniques may, however, perform poorly in highly irregular settings. Figure 3 shows nearest neighbor interpolation with the Dimensions storage and search system [15] in a highly irregular network. Briefly, the system stores wavelet summaries at different resolutions, and uses these summaries to route queries to parts of the network that are most likely to provide a good answer. We do not go into significant detail on the system for lack of space, but point out that, as shown in Figuree 3, the performance of a simple MAX query is poor in the irregular case.

These results reflect the need for more sophisticated spatial processing including higher order interpolation, especially if node distribution is highly skewed as in our example. Such techniques, however, involves more communication than in the nearest-neighbor case, since samples from a larger area is required for better interpolation.

**Time-series alignment**

In the previous section, we addressed how a junction node on a tree can deal with irregular spatial sampling. As we mentioned earlier, loosely synchronized clocks lead to poor time-series compression as well.

Figure 4 shows two difficulties in dealing with loose sychronization and clock variability due to ambient environmental conditions. Loose synchronization results in *lag* between the two time-sequences, which similar to phase lags, can be corrected using the commonly used method of cross-correlation. Wang et al [19] describe a computationally efficient manner for estimating the lag between time-series signals of bird-calls that uses cross-correlations. However, as shown in Figure 4, the lag between time-series is not constant with time, in fact, for a pair of ipaqs detecting a stationary source, the lag doubles within a couple of seconds. This results from variations in clock frequency, which can be expected when sensors in different parts of the network are subject to different conditions (such as one sensor in shade and the other in the sun). Some simple techniques such as chopping the time-series into segments and assuming fixed lag within each segment can be an initial heuristic to solve this problem. However, the problem raises many difficult questions that future research will need to consider.

# 3 Impact of spatial irregularity on data storage

Energy-efficient data-gathering techniques are required to extract relevant sensed data from within a sensor network. Early work [5, 7] relied on a flood-and-respond approach in which queries are flooded and responses (data) are routed back to the querying node. More recently, in-network storage has been proposed as an alternative approach to the data-gathering problem. Under this approach, data are stored by name at nodes within the sensor network; all data with the same general name are thus stored at the same node (typically not the one that originally gathered the data). Queries for data with a particular name can then be sent directly to the node storing those named data, thereby avoiding the query flooding typically required in previous approaches.

Recent research has seen a a growing body of work on data storage schemes for sensor networks [20, 15, 21]. These techniques differ in the aggregation mechanisms used, but are loosely based on the idea of geographic hashing and structured replication. In geographic hashing, names are hashed to points (locations) in the geographic space occupied by the sensor network. Data is then stored at the node closest to the location obtained by hashing its associated name. GPSR is used as the routing mechanism to reach the node closest to a target location. Structured replication [20] uses a hierarchical decomposition of the geographic space into nested grids that is useful for aggregation and lends locality in storage.

These two underlying primitives – geographic hashing and structured replication – build on two assumptions: (a) all nodes know the external boundary of the sensor network, and (b) distribution of nodes is uniform over the entire topology. For example, in GHT, the location at which data is stored is determined by uniformly hashing names over the geographic region occupied by the sensor network, which is assumed known to every node.

Two problems arise with irregular topologies where the true geographic extent does not match that used for hashing, and the node distribution is non-uniform:

**Skewed storage load:** All data hashed to points outside the true external perimeter or to a sparsely deployed region in the network are stored at a relatively small number of perimeter nodes, thus greatly skewing the storage load. For instance, in Figure 1(a), the upper left region is sparsely deployed, and the lower right region is outside the external network boundary. Hashing to either of these regions would result in skewed storage load.

**High routing overhead:** To reach the node closest to a hashed location, GPSR traverses the perimeter enclosing the target location before terminating appropriately at the closest node. For target locations outside the external perimeter, GPSR's perimeter mode will traverse the *entire* external perimeter of the network before terminating at the appropriate storage node. For example, in Figure 1(a): any attempt to store data at positions south of sensor 36 and east of sensor 25 would end up traversing the entire boundary of the network. For the same reasons, hashing to a sparse region (such as the upper left of the figure) would result in high perimeter forwarding overhead, however, the impact is less severe.

## 3.1 Coping with irregularity in data-centric storage

As described above, to accomodate irregular topologies, data-centric storage schemes must address two issues: (a) discovery of the network's external boundary and (b) dealing with internal variations in node density. We now briefly describe two very different approaches to dealing with the former and some simple heuristics that address the latter.

**Boundary tracing:** Boundary estimation algorithms such as [14] can be used to discover the true external perimeter of the network, which is then disseminated to all nodes. Such a procedure has two potential drawbacks: (a) it is expensive since boundary representations have to be communicated to *every* node in the network, and (b) the boundary tracing process would need to be periodically repeated to accomodate system dynamics (such as node failures or node mobility). Distributing coarse boundary estimates (eg: constructed using wavelets [15]), rather than more precise ones, can potentially mitigate the above overhead.

**Virtual Coordinate Spaces:** A different approach, enabled by recent work on scalable ad-hoc routing algorithms (Rao et al. [22] and Song et al. [23]), is to embed nodes into a well-known "virtual" region. These virtual coordinates need not be accurate representations of the underlying geography but, in order to serve as the basis of routing, must capture the underlying *relative* connectivity between nodes. Using these algorithms, in-network storage can be achieved by hashing names over the virtual coordinate space instead of the geographic space as in GHT [22]. This approach to in-network storage deals more easily with irregular topologies because the extent of the virtual coordinate space is pre-defined and well known, thus eliminating the need to discover the external perimeter of the network.

**Heuristics for local adaptation to irregular node densities:** Within a network's boundaries, simple local heuristics that dynamically adapt to local irregularities in the placement of nodes may be effective for small irregularities. For example, nodes can discover local voids in their immediate vicinity and any data destined to locations within such voids are rehashed and routed to an alternate destination. Yet another option might be to use multiple hash functions to store multiple copies of the data (such replication is probably needed for availability reasons in any case) and to abort a store/retrieve operation if its perimeter traversal requires more than a threshold number of hops.

# 4 Discussion and Conclusions

In this paper, we argued that spatio-temporal irregularity is fundamental to wireless sensor networks. Many factors contribute to this irregularity: application sampling requirements, difficulty in deploying nodes on the terrain, availability of power sources and radio connectivity, variability in clock synchronization, etc. We focused on two research areas: data compression and data storage, and suggested a spectrum of schemes to deal with the challenges of irregularity.

It is likely that irregular spatio-temporal sampling will have major implications for many other key research topics in sensor networks as well. For example, a widely researched topic in both ad-hoc and wireless sensor networks is the problem of creating well-connected topologies. Experimental results (eg: [24]) have shown that there are significant differences between communication cells in a real channel and a unit disk model. While percolation theory for regular networks have shown sharp transitions in connectivity when unit disk models are used, there is a growing body of literature that indicates that communication irregularities may result in dramatic differences, including non-existence of such thresholds in some extreme cases ([25]). Thus, the impact of irregularity cuts across research themes in sensor networks, and warrants deeper examination of existing results and techniques.

# References


[1] Michael Hamilton. James San Jacinto Mountains Reserve.

[2] A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, and J. Anderson. Wireless sensor networks for habitat monitoring. In *ACM WSNA*, Atlanta, GA, September 2002.

[3] S. Ganeriwal, C.C. Han, and M.B. Srivastava. Going beyond nodal aggregates : Spatial average of a continuous physical process in sensor networks. Poster in ACM Sensys 2003, to appear.

[4] G.J. Pottie and W.J. Kaiser. Wireless integrated network sensors. *Communications of the ACM*, 43(5):51–58, May 2000.

[5] C. Intanagonwiwat, R. Govindan, and D. Estrin. Directed diffusion: A scalable and robust communication paradigm for sensor networks. In *ACM/IEEE Mobicom*, Boston, MA, 2000.

[6] J.M. Kahn, R.H. Katz, and K.S.J. Pister. Next century challenges: mobile networking for Smart Dust. In *ACM/IEEE Mobicom*, 1999.

[7] S. Madden, M. Franklin, J. Hellerstein, and W. Hong. Tag: a tiny aggregation service for ad-hoc sensor networks. In *OSDI*, Boston, MA, 2002.

[8] A. Cerpa and D. Estrin. Ascent: Adaptive self-configuring sEnsor networks topologies. In *IEEE Infocom*, 2002.

[9] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris. Span: An energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks. In *ACM Mobicom*, 2001.

[10] S. Ratnasamy, D. Estrin, R. Govindan, B. Karp, L. Yin S. Shenker, and F. Yu. Data-centric storage in sensornets. In *ACM Hotnets*, 2001.

[11] Jeremy Elson, Lewis Girod, and Deborah Estrin. Fine-grained network time synchronization using reference broadcasts. In *OSDI*, Boston, MA, 2002.

[12] B. Karp and H. T. Kung. GPSR: greedy perimeter stateless routing for wireless networks. In *Proceedings of Mobicom*, 2000.

[13] Richard Karp, Jeremy Elson, Deborah Estrin, and Scott Shenker. Optimal and global time synchronization in sensornets. Technical report. CENS TR-012, UCLA.

[14] R. Nowak and U. Mitra. Boundary estimation in sensor networks: Theory and methods. In *IPSN '03*, 2003.

[15] D. Ganesan, B. Greenstein, D. Perelyubskiy, D. Estrin, and J. Heidemann. An evaluation of multi-resolution storage for sensor networks. In *ACM SenSys.*, 2003. to appear.

[16] S. Pradhan, J. Kusuma, and K. Ramchandran. Distributed compression in a dense sensor network. *IEEE Signal Processing Magazine*, 1, March 2002.

[17] S. D. Servetto. Sensing lena—massively distributed compression of sensor images. In *Proceedings of the IEEE ICIP*, 2003.

[18] E. Meijering. Chronology of interpolation: From ancient astronomy to modern signal and image processing. In *IEEE*, volume 90, 2002.

[19] H. Wang, J. Elson, L. Girod, D. Estrin, and K. Yao. Target classification and localization in habitat monitoring. In *ICASSP*, volume 1, 2003.

[20] S. Ratnasamy, B. Karp, L. Yin, F. Yu, D. Estrin, R. Govindan, and S. Shenker. Ght - a geographic hash-table for data-centric storage. In *ACM WSNA*, 2002.

[21] X. Li, Y.J Kim, R. Govindan, and W. Hong. Multi-dimensional range queries in sensor networks. In *ACM SenSys.*, 2003. to appear.

[22] A. Rao, C. Papadimitriou, S. Ratnasamy, S. Shenker, and I. Stoica. Geographic routing without location information. In *ACM Mobicom*, 2003.

[23] J. Newsome and D. Song. Gem: Graph embedding for routing and data-centric storage in sensor networks without geographic information. In *ACM SenSys.*, 2003. to appear.

[24] D. Ganesan, B. Krishnamachari, A. Woo, D. Culler, D. Estrin, and S. Wicker. Complex behavior at scale: An experimental study of low-power wireless sensor networks. Technical Report UCLA/CSD-TR 02-0013, UCLA, Department of Computer Science, 2002.

[25] Massimo Franceschetti et al. Percolation in multihop wireless networks. Under Submission.