Networking Wireless Sensors: New Challenges

Bhaskar Krishnamachari
Autonomous Networks Research Group
Department of Electrical Engineering-Systems
USC Viterbi School of Engineering
http://ceng.usc.edu/~anrg
bkrishna@usc.edu
Large numbers of resource constrained wireless devices, providing a high resolution spatio-temporal interface between the physical and virtual worlds.

Applications include industrial process monitoring, environmental sensing, ecological studies, military surveillance, structural monitoring, etc. Considerable academic and industry interest in this field in recent years.

These systems present a number of unique challenges and novel research problems that are often fundamentally different from those encountered in traditional networks.
Challenges

- Unattended ad-hoc deployment
- Very large scale
- Scarce energy and bandwidth resources
- High noise and fault rates
- Dynamic / uncertain environments
- High variation in application-specific requirements
Case Studies

- Phase Transition Phenomena
- Robust, Energy-Efficient Geographic Routing
- Impact of Spatial Data Correlation on Routing with Compression
- Delay Efficient Sleep Scheduling
1. Phase Transitions

Illustration: Connectivity in an randomly deployed wireless network
Phase Transition for Connectivity

Energy-efficient operating point

- Undesirable regime
- Desirable regime

Communication radius $R$

Probability (Network is Connected)

$n = 20$
Phase Transition for Connectivity

Communication radius $R$

Probability (Network is Connected)

sharp transition with increasing $n$

![Graph showing probability of network connectivity with increasing communication radius $R$.]
Bernoulli Random Graphs $G(n,p)$

- Studied by mathematicians since 60’s (Renyi, Erdos, Bollobas)
- A number graph properties, including connectivity, appear abruptly at a critical value of $p$ which is the density parameter
- In 1996, Friedgut and Kalai proved a fundamental result: all monotone properties in Bernoulli Random Graphs show sharp phase transitions from 0 to 1
Geometric Random Graphs $G(n, R)$

- A good model for Wireless Ad Hoc / Sensor Networks
- Density Parameter: radius $R$. Captures locality of communication and sensing.
- Gupta and Kumar showed in 1998 that connectivity has a sharp threshold in such graphs (implied also in earlier independent work by Penrose).
- Triangle inequality affects independence, so different proof-techniques needed for such graphs.
Beyond Connectivity: Hamiltonian Cycle

Beyond Connectivity: Sensor Tracking

Can all targets be tracked by three communicating sensors?

**Conjecture:** An analog of the Friedgut-Kalai result exists for geometric random graphs.
New Result: Sharp thresholds exist for all monotone properties in G(n,R)

- Formally, let the threshold width $\delta(n, \varepsilon)$ of a property in a geometric random graph be the difference in radius between when the property is satisfied with probability $1-\varepsilon$ and probability $\varepsilon$. The Theorem states that the threshold width of any monotone property $\delta(n, \varepsilon) = o(1)$ (i.e., asymptotically tends to zero). Specifically, it is proved that $\delta(n, \varepsilon) = O(\log^{3/4}n/\sqrt{n})$ (for $d = 2$), and $O(\log^{1/d}n n^{-1/d})$ for $d > 2$.

- The proof involves the use of bounds on bottleneck matching lengths in random geometric graphs. A corollary result shows that a random geometric graph is a subgraph (w.h.p) of another independently drawn random geometric graph with a slightly larger radius.

- Thus, sharp density thresholds exist not only for connectivity but for all other graph properties of interest in sensor networks - including k-connectivity, k-coloring, Hamiltonian cycles, sensor coverage, etc.

2. Geographic Routing over Wireless Links

- Network protocols designed with binary links in mind may show really poor performance: e.g. Greedy Geographic Routing (forward to neighbor showing best distance improvement).
Real Wireless Links

- Low-power wireless channels are harsh and unreliable - not binary links.

Fig. 11. Empirical Validation of the Model for the Outdoor Channel. (a) and (b) are the empirical measurements, and (c) and (d) their analytical counterparts.

Two Extremes

1. Forward to best-distance improvement neighbor
   - Pro: fewer total hops
   - Con: each long hop likely to have low PRR, hence, may need many retries to get packet across each hop.

2. Forward to nearby neighbor in direction of destination
   - Pro: each hop is likely to be high PRR, hence energy-efficient
   - Con: Requires more total hops, since only a short distance is traveled at each step

• What’s the right compromise?
The PRR*D metric

- Optimal Solution: in between these extremes. We have shown that the best local metric (in terms of packets delivered per unit energy) for geographic forwarding in a multi-hop wireless network is the product of link packet reception rate and distance improvement.
3. Impact of Spatial Correlation on Routing with Compression

- Consider a scenario where data is being gathered from a sensor network monitoring a physical environment.

- It is now understood that significant energy gains can be obtained by combining routing with in-network compression, but what is not well understood is the impact of different levels of spatial correlation on joint routing and compression.

- To quantify the energy costs of communication, we use the joint entropy of n sources to indicate the total amount of compressed information generated by them.

Spatial Correlation Model

A parameterized expression for the joint entropy of n linearly placed equally spaced nodes:

\[ H_n(d) = H_1 + (n - 1) \left[ 1 - \frac{1}{\left( \frac{d}{c} + 1 \right)} \right] H_1 \]

- Entropy of single source \( H_1 \)
- Number of nodes \( n \)
- Inter-node spacing \( d \)
- Correlation level \( c \)
- [c=25, RMS error = 0.09]
- [c=25, RMS error = 0.055]
- [c=25, RMS error = 0.03]
Comparison of Basic Strategies

- **Routing Driven Compression**: route along shortest paths to sink, compress wherever paths happen to overlap.

- **Compression Driven Routing**: Route to maximize compression, though this may incur longer paths.

- **Distributed Source Coding (ideal)**: perform distributed compression at sources, and route along shortest paths. If we ignore costs of learning correlation, this provides an idealized lower bound.
Comparison of Basic Strategies

![Graph showing energy usage in bits/ochs against correlation parameter in log scale log(c). Dashed line represents CDR, solid line represents RDC, and dotted line represents DSC.](image)
Generalizing the strategies

- We find that a CDR strategy works well with high correlation, and an RDC strategy works best at low correlation -- what about in between?

- We can look for some hybrid techniques, but there must be some systematic way to parameterize them, if we are to have any hope of analysis.

- Solution: Clustering. Create clusters of nearby sources. First route to compress the data within the sources inside each cluster, then route the compressed data from the clusters towards the sink along shortest paths and compress further where there is overlap. RDC and CDR are then the two extreme special cases (cluster of 1 source, and cluster of n sources).
Analysis

- Consider again the linear set of sources, in a 2D grid.

- We can derive expressions for the energy cost as a function of the cluster size $s$:

$$E_s(c) = n H_1 [1 + \frac{(s - 1)}{2(1 + c)} + \frac{D}{s} + \frac{(s - 1)D}{(s)(1 + c)}]$$

- We can even derive an expression for the optimal cluster size as a function of the network size and correlation level:

$$s_{opt} = \sqrt{2Dc}$$
Cluster-based routing + compression

Suggests the existence of a near-optimal cluster (about 15) that is insensitive to correlation level!
Near-Optimal Clustering

- Can formalize the notion of near-optimality using a maximum difference metric:

\[
\min_s \max_{c \in [0, \infty)} \left| E_s(c) - E^*(c) \right|
\]

- We can then derive an expression for the near-optimal cluster size:

\[
s_{no} = \frac{\sqrt{8n + 1} - 1}{2}
\]

- This is *independent* of the correlation level, but does depend on the network size, number of sources, and location of the sink. For the above scenario, it turns out \( s_{no} = 14 \) (which explains the results shown).
Near-Optimal Clustering

4. Delay-Efficient Sleep Scheduling

- Largest source of energy consumption is keeping the radio on (even if idle). Particularly wasteful in low-data-rate applications.

- Solution: regular duty-cycled sleep-wakeup cycles.

- Say there are $k$ slots, each node stays awake for one of these slots to receive packets intended for it. It makes this cycle known to its neighbors so they know when to wake up and transmit a packet to it.

- While this limits energy consumption, it can introduce significant sleep delay. E.g. if all nodes synchronize to wake and sleep at the same time, then a packet will take $m$ cycles ($m \times k$ slots) to traverse $m$ hops in the network.
Fig. 1. Examples of slot assignment with $k = 3$. The dotted arrows show the delay on each link in the corresponding direction.
Problem Formulation

- DESS: Given a graph $G$, assign from $k$ slots to minimize the maximum delay between any two points in the network.

- The DESS problem is NP-hard.

- Some provable polynomial special cases:
  - Ring: sequential slot assignment has best possible delay diameter of $d$.
  - Tree: alternate between 0 and $k/2$. Gives worst delay diameter of $dk/2$ ($d$ is the graph diameter).

DESS on a grid

- Assignments (believed to be near-optimal) obtained using Simulated Annealing show interesting structure, but no obvious general approach.

Fig. 7. A slot assignment $f$ obtained for (a) a $5 \times 5$ grid with $k = 4$ and $D_f = 9$ (b) $5 \times 5$ grid with $k = 16$ and $D_f = 20$ using simulated annealing. The dotted arrows show the sequential assignment on cycles in the grid.
Open Problem

- No known bounds or provably efficient solutions/approximations exist for this problem. Distributed solutions seem to show particularly poor performance.

Fig. 10. The delay diameter of the heuristic algorithms versus the number of slots ($k$) for a fixed grid size of 20x20.
Summary

• We examined a number of interesting case studies
  – existence of phase transitions in geometric random graphs,
  – existence of near-optimal routing+compression structures,
  – metric for efficient geo-routing in the face of unreliable links
  – delay efficient radio sleep schedules

• They show the novelty and richness of the design and analytical challenges involved in networking wireless sensors.
Sequential Paging in Cellular Wireless Networks

- Need to locate mobile user upon call arrival

- 3 step process:
  - mobile provides location updates upon change in location area
  - network pages **all cells** in location area upon call arrival
  - mobile responds to page

- this control traffic can incur very high radio bandwidth costs, particularly in high-capacity, high-traffic cellular wireless networks.
Sequential paging

- Goal: minimize number of paging messages by using probabilistic location estimates. Example:
  - Average Paging Cost $E[L]$: $0.6 \times 2 + 0.4 \times 5 = 3.2$ pages
  - Worst Case Delay $w$: 2 rounds
  - Average Delay $E[D] = 0.6 \times 1 + 0.4 \times 2 = 1.4$ rounds

<table>
<thead>
<tr>
<th>Cell</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>prob.</td>
<td>0.3</td>
<td>0.3</td>
<td>0.25</td>
<td>0.1</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Results

• Problem of minimizing average paging cost subject to a constraint on the average delay was believed to be intractable:
  • “This problem is not amenable to solution via Dynamic Programming owing to the constraint on E[D]” - Rose & Yates 1995.
  • “(This Problem) is NP-complete” - Abutaleb & Li, 1997

• We developed an $O(n^3)$ dynamic programming solution to this problem, proving its tractability.

• Also derived analytical results quantifying the performance of sequential paging, showing that very high gains are possible when the user location probabilities are concentrated within a small subset of the full region.

• Open Question: how can these probabilities be estimated in practice?

Integrating Cellular & Sensor Networks

- Compelling reasons to look at the intersection of these two domains, since cellular networks represent an already-existing large-scale wireless system.

- Questions:
  - Can the cellular network itself be treated as a sensor network, by equipping cellular devices with sensing capabilities?
  - Can the cellular network infrastructure be leveraged to extract data from large scale pervasive sensor systems (e.g. by designing mechanisms to route to random, mobile sinks)?

- These questions are being pursued already in commercial settings, notably in the Pervasive Indeterminate Measurement Systems (PIMS) work by researchers at Agilent (J. Warrior, J. Eidson et al.).