Title
BARD: Bayesian Assisted Resource Discovery

Authors
Fred Stann
John Heidemann

Publication Date
2004

Center for Embedded Networked Sensing

BARD: Bayesian Assisted Resource Discovery

Fred Stann, John Heidemann
fstann@usc.edu, johnh@isi.edu, http://www.isi.edu/ilense

USC VITERBI SCHOOL OF ENGINEERING

ISI Information Sciences Institute

I-LENSE

CENS

Problem Description: Flooding Limitation with Attribute-Based Data Dissemination

Efficient Data Resource Discovery

- **Goal: Flooding Limitation**
  - Without a priori knowledge, location of data resources/sinks is expensive
    - Flooding in Directed Diffusion
    - Even non-queried data is constantly moved with Data Centric Storage
  - Efficient data resource discovery should be available to applications independent of type or number of data attributes
    - Flooding-limitation schemes rely on specific attributes (e.g., geography)
- **Exploit Spatial Correlation**
  - Data can be spatially correlated by a large variety of problem aspects: topography, target tendencies, geography, sink locations, link reliability, inactive regions, etc…

- **The scenario**
  - Some nodes (red) will not lead to successful data dissemination paths for either sink, other nodes (green) will generate data for both sinks, and some (black) are helpful to only the diag sink
  - Relative position of active sources and sinks dictates how many nodes will be involved in reinforced data paths
  - Links that historically (over time) contribute to source-to-sink paths define an "implicit geography"

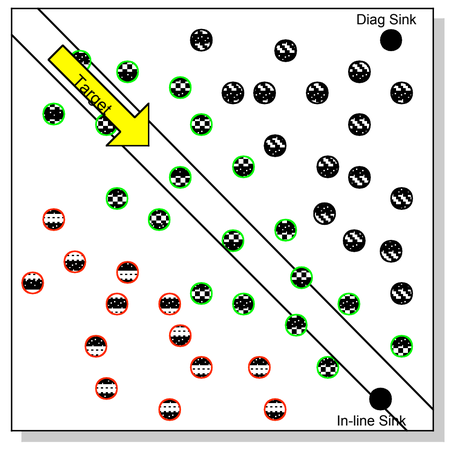


Figure 1
Fifty node sensor network intended to track target movement, where targets display spatial correlation. Two alternate sink locations (diagonal vs. inline) are depicted

Approach: Per-attribute history of data paths as a statistical predictor of future success

Probabilistic Approach

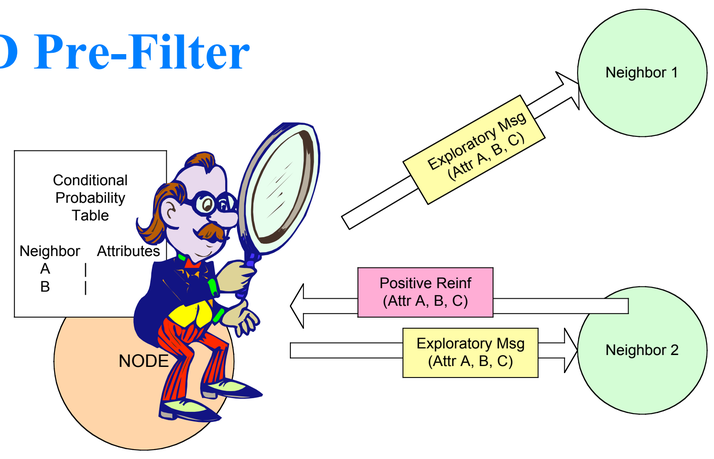
- Use prior routing history to avoid flooding for similar queries
- Collect per-attribute statistics about which neighbors lead to data resources in a sliding window of time
- Use the collected statistics (prior distribution) to limit flooding for 4/5ths of subsequent queries
- Use occasional flooding ($1/5^{th}$) to maintain prior distribution

Bayesian Estimation

- *Naïve Bayes* can build probabilities incrementally from the set of attributes that defines a query or data
- Even partial matches contribute to probability calculation
- Limit flooding to high probability candidates (neighbors)
- Fully distributed scheme without inter-node transfer of meta-data

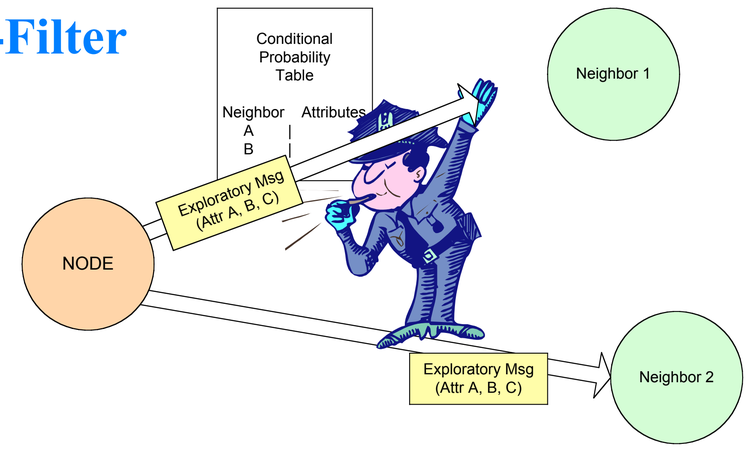
Proposed Solution: Use Bayesian Estimation to Limit Flooding in Diffusion

BARD Pre-Filter



- **Establish the prior distribution**
  - Examination of diffusion reinforcements establish success
  - Statistics kept on a per-attribute + per-neighbor basis
  - $P[N_1|A \cap B \cap C] = \alpha\, P[N_1]\, P[A|N_1]\, P[B|N_1]\, P[C|N_1]$

BARD Post-Filter



- **Limit Flooding**
  - Build probabilities incrementally from per-attribute CPTs
  - Squelch flooding to low-probability neighbors

Results

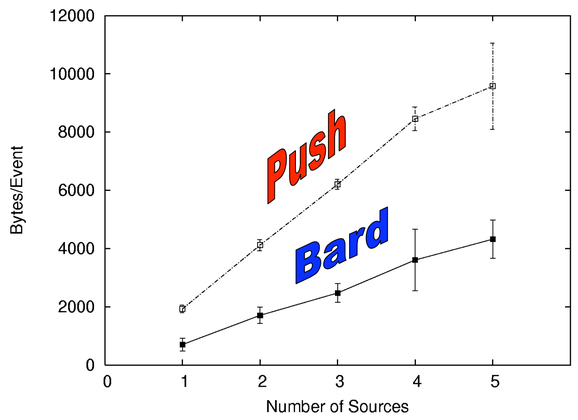


Figure 2
Normalized control byte transmissions per event required to transfer 90 packets of 128 bytes each from a variable number of randomly placed sources to a single sink.

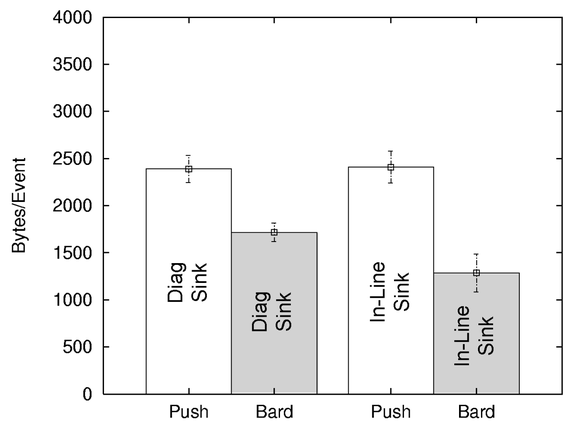


Figure 3
Control byte overhead for target tracking application (in Figure 1). BARD routing vs. simple Push when sink is diagonal from target route vs. in-line with target.

- **What improvement does BARD show over Push diffusion with 50 randomly placed nodes, average node density of 10.9 & varying number of sources?**
  - BARD must initially "learn" where the sources are
  - Percentage of improvement decreases from 63% to 54%
  - As more of the topology is involved in routing the degree of flooding limitation is necessarily less.

- **How does spatial correlation affect efficacy of BARD ?**
  - A target moves along a similar same path repeatedly
  - BARD "learns" the path and limits flooding to upper diagonal or in-line path
  - Improvement by BARD is 28% for diagonal sink and 45% in-line
  - Percentage of improvement by BARD is dependent on degree of fan-out of routing tree from sink.

Conclusions

- Improvement from BARD is in proportion to spatial correlation of events in environment and size (fan-out) of routing tree
- BARD is a general technique applicable to a wide variety of applications with arbitrary attributes for their named data
- Bayesian estimation allows for partial matches where there is still a high probability of data and query matching
- Because it is implemented as a diffusion filter it is a selectable service to users of diffusion

BARD paper available at:
http://www.isi.edu/~fstann/papers/ISI-TR-2004-593.pdf