SMITE: A Stochastic Compressive Data Collection Protocol for Mobile Wireless Sensor Networks

Longjiang Guo, Raheem Beyah, and Yingshu Li

Department of Computer Science, Georgia State University, USA

Outline

- Application Background
- Problem Description
- Faced challenges
- Existing work
- SMITE
  - Core idea
  - Other aspects of SMITE
  - Probability analysis
- Simulation Evaluation
- Conclusion
Collars with wireless sensors are attached to zebras’ necks to study their habitat and to collect environmental data.

**ZebraNet and Beyond: Applications and Systems**
**Support for Mobile, Dynamic Networks**
Dept. of Electrical Engineering, Princeton University

April 10-15, 2011 infocom
data collection in mobile WSNs

ZebraNet vs. Other SensorNets
- All sensing nodes are mobile
- Large area: 100’s-1000s sq. kilometers
- GPS on-board
- Long-running

How to collect each data on every node in this random mobile wireless sensor network?

Data is very important for scientists.
Problem Description

- Data is very important for scientists.
- The research problem in this paper is:
  - Design a data collection protocol in mobile wireless sensor networks to collect all data on each node based on random waypoint mobility.
  - Objectives:
    - Low energy consumption
    - High data delivery ratio
    - Low time delay

Problem Description

- This paper considers a mobile WSN consisting of one mobile sink and \( N \) mobile sensor nodes randomly moving in an area of \( Z \times Z \), where \( Z \) is the width of the monitored area.
- The sensor nodes and the sink set their radio transmission range as \( r \).
- Our assumption:
  - Nodes are synchronized and each node knows its velocity as well as its position using GPS.
Faced challenges of Gathering data from mobile sensors

- Traditional challenges in static WSN
  - Limited energy
  - Narrow bandwidth
  - Limited computing ability

- Volatile topology
  - The connectivity of the network constantly changes since sensors always move.
  - The topology is dynamic and volatile at all times.
  - Network topology of MWSNs changes more acutely than that of mobile adhoc networks, because sensor radios have a much smaller radio range
    - Outdoor range 75m ~100m for TelosB TPR2400CA compared with mobile ad-hoc networks 150m ~ 250m in 802.11.

- Loose connectivity
  - A mobile sensor is connected to other sensors only randomly and occasionally due to the nodes’ mobility.
Faced challenges of Gathering data from mobile sensors

- **High message overhead**
  - Each node needs to monitor a large set of parameters (*multi-dimensional data*), which makes the data size large. A MWSN could consist of hundreds to thousands of sensors. Thus, tremendous amounts of data would be generated and delivered to the sink.
  - Communication messages for data collection protocol is also indispensable.

- **Limited buffer space**
  - Each sensor node has a amount of limited buffer space. This constraint has a significant impact on MWSNs, because sensor nodes store the received data locally prior to relaying the data.
  - Limited buffer space makes it impossible to store a large amount of data locally. Consequently, some useful data is lost and more subsequent communications are incurred.
Related work

- Conventional routing protocols [5](MOBICOM, 2009, C. Luo et al) [7](SENSYS, 2003, A. Woo et al) designed for static WSNs are not suitable for MWSNs.

- Tree-based routing protocol [6] (ICDCS, 2010, Y. Li and L. Guo), may be effective for slight topology changes caused by nodes’ sleeping and failures, however it still not robust enough to conquer the volatile topology.

- Geographic forwarding-based routing protocols [8] (GPSR, MOBICOM, 2000, B. Karp et al) [9](MOBIHOC, 2006, P.-J. Wan et al) [10] (ICDCS, 2003, SPEED, T. He et al) are also not applicable for MWSNs because intermediate nodes do not know where a mobile sink is during the process of forwarding messages.


- Through extensive experiments,[14](Sidewinder, SECON, 2009, M. Keally et al) have validated that these protocols cannot be applied to MWSNs
  - These protocols need to establish a routing path from a source node to the sink before data is sent.
  - This strategy may work well for slight topology changes.
  - It is not a good choice for volatile topology.
Related work

- Some existing works [15](INFOCOM 2010, X. Xu et al)[16](SenCar, TPDS, 2007, M. Ma, et al) utilize mobile sinks to collect data in static WSNs.
  - However, these works only consider the case where sinks are mobile and sensor nodes are static.

- [17](VTC, 2006, X. Zhang et al),
  - Each node dynamically chooses its cluster-head based on its current velocity to facilitate data transmission to the nearest cluster-head.
  - A cluster-head is responsible for gathering data and forwarding data to the static sink through a single hop.
  - This assumption is not reasonable, especially for large-scale greatly-dynamic MWSNs where a cluster-head has to send its data to the base station through multiple hops considering the fact that the radio range in MWSNs is small.
Most related work

- **DFT-MSN** (INFOCOM, 2006, Y. Wang et al)
  - Data is forwarded to mobile sink by optimized flooding.

- **SCAR** (MASS, 2007, B. Psztor et al)
  - Data is forwarded to mobile sink by finding an appropriate neighbor.
  - To do this, each node should periodically broadcast its info including speed, distance to sink, and locations.

- **Sidewinder** (SECON, 2009, M. Keally et al)
  - Sink communicates with its neighbors for computing sink’s group’s average location and speed.
  - Each node if receives a message from sink, it should broadcast its info including speed, distance to sink, and locations.
  - It is more suitable for group mobility model.

- These approaches have high message overhead.

Contribution of this paper

- SMITE transforms the traditional data collection problem in MWSNs into in-network data aggregation through a data compression technique using bloom filters.

- SMITE is completely different from the traditional idea where every node periodically broadcasts its context information to its neighbors to maintain real context. In SMITE, only collectors periodically broadcast their context information.

- The analytic results show that data on the common nodes can be transmitted to collectors with a high probability and complete data gathered on collectors can be further forwarded to the mobile sink with a high probability using angle transmission.
Core idea of SMITE

- SMITE has three phases:
  - The first phase is collector election.
  - The second phase is direct data transmission from non-collectors to collectors while non-collectors are in the communication range of collectors.
  - The third phase is data forwarding from the collectors to the mobile sink using angle transmission.
Core idea of SMITE

April 10-15, 2011

Collector Node 7

(b)

Core idea of SMITE

April 10-1

(c)
Core idea of SMITE

\[ \dot{r}_u = v(t_x)(t_{now} - t_k) \]

Sink's movement territory

(d)

Core idea of SMITE

(e)
Core idea of SMITE

Estimation for sink’s movement territory

For any relayed node $w$, it maintains the recent $k$ sink’s velocities $v_{t_1}, v_{t_2}, \ldots, v_{t_k}$ and sink’s recent position $(x_{t_k}, y_{t_k})$

- From $t_k$ to current time slot $t_{now}$, the sink’s movement distance can be calculated according to Integral Mean Value Theorem as follows:

$$\int_{t_k}^{t_{now}} v(t) dt = v(t_\xi) (t_{now} - t_k), t_\xi \in [t_k, t_{now}]$$
Estimation for $v(t_\xi) \quad t_\xi \in [t_k, t_{\text{now}}]$

the recent $k$ sink’s velocities $v_{t_1}, v_{t_2}, \ldots, v_{t_k}$

sink’s recent position $(x_{t_k}, y_{t_k})$

Second difference transformation

\[
\begin{align*}
W_{t_1} &= v_{t_1} - v_{t_{i-1}} - (v_{t_{i-1}} - v_{t_{i-2}}); (i = 1, 2, \ldots, k) \\
W_{t_1} &= v_{t_1} v(t_\xi)(t_{\text{now}} - t_k); t_\xi \in [t_k, t_{\text{now}}]
\end{align*}
\]

Weighted moving average as predict model

\[
\begin{align*}
\hat{W}_{t_1} &= W_{t_1} ; \hat{W}_{t_2} = W_{t_2} ; \cdots ; \hat{W}_{t_k} = W_{t_k} \\
\hat{W}_{t_i} &= (1 - \lambda) \lambda^{k-1} \hat{W}_{t_{i-k}} + \cdots + (1 - \lambda) \lambda^{k-i} \hat{W}_{t_{i-1}} \\
i &= k+1, k+2, \ldots, \xi.
\end{align*}
\]

$\lambda$ is a system parameter from $[0, 1]$. $t_\xi \in [t_k, t_{\text{now}}]$

$\hat{v}_{t_i} = \hat{W}_{t_i} - \hat{v}_{t_{i-2}} + 2\hat{v}_{t_{i-1}}(i-k+1, k+2, \ldots, \xi)$.

---

How can node $w$ determine whether it is in the angle $\angle AB$ while $w$ receives the packet from node $u$?

![Diagram](image)

Fig. 3. Prediction of the sink’s movement territory.

\[
\begin{align*}
|u||w| &\leq r \\
\arccos\left(\frac{|u||w|^2 + |u||w|^2 - |w||w|^2}{2|u||w||w|}\right) &\leq \arcsin\left(\frac{\hat{r}_u}{|u||w|}\right)
\end{align*}
\]
Other aspects of SMITE

- Sink’s partial flooding
- Data Compression and Buffer Management
- Rotation of The Collectors

Sink’s partial flooding

$h_w$ indicates the location correlation between a relayed node and the sink.

$$h_w = \sqrt{(x_w - x_t)^2 + (y_w - y_t)^2} / r$$

$(x_w, y_w)$ denotes $w$’s position
$r$ is the communication range

If $h_w < h_{w} \leq h_{max}$ and received VP is a new message, then $w$ continues to broadcast
Otherwise, $w$ drops the received packet

According to Corollary 8

$$h_{max} = \sqrt{N \cdot p}$$

Fig. 2. The location correlation between a relayed node and the sink.

For any relayed node $w$, it maintains the recent $k$ sink’s velocity information $v_{t_1}, v_{t_2}, \ldots, v_{t_k}$, and the most recent sink’s position $(x_{t_k}, y_{t_k})$, where $t_k$ is the most recent time
Sink’s partial flooding

**Corollary 8**: For any node $w$, the probability that $w$ receives the $VP$ message from the mobile sink during $\frac{1}{1-c}$ time slots is $\frac{h_{\text{max}}^2}{Np}$, where $h_{\text{max}}$ is the number of the flooding hops from the sink.

$$ b = \frac{\pi r^2}{Z^2} \text{Pr}(X_{w,t} = 0) = (1 - b)^{(N \cdot p)} = c $$

$$ h_{\text{max}} = \sqrt{N \cdot p} $$

---

**Data Compression and Buffer Management**

$$(c_1, c_2, \cdots, c_k)$$

$$cm < \sum_{i=1}^{k-1} (c_i \cdot \prod_{j=i+1}^{k} D_j) + c_k$$

$$(p,q) \leftarrow \log_2(cm); \quad p \text{ is the integral part}$$

$$m_w \leftarrow q \cdot D_I + p$$

$C_i$’s domain size is $D_1$
**Un-compression**

\[ p \leftarrow m_v - \left[ \frac{m_v}{D_I} \right] \cdot D_I; \quad q \leftarrow \frac{m_v-p}{D_I}; \quad cm \leftarrow 2^{(p-q)}; \]
\[ cm_k \leftarrow cm; \quad c_k \leftarrow cm_k - \left[ \frac{cm_k}{D_k} \right] \cdot D_k; \]
\[ cm_{k-1} \leftarrow \frac{cm_k-c_k}{D_k}; \quad c_{k-1} \leftarrow cm_{k-1} - \left[ \frac{cm_{k-1}}{D_{k-1}} \right] \cdot D_{k-1}; \]
\[ \cdots \]
\[ cm_1 \leftarrow \frac{cm_2-c_2}{D_2}; \quad c_1 \leftarrow cm_1; \]

---

**Data storage using BF**

\[ x = \frac{N-N_p}{N_p} + 1 = \frac{1}{p} \]
\[ BF(L, k, \eta, \min, \max, n) \]
\[ L = 16 \cdot x = \frac{16}{p} \]
\[ f_{L,k,n}^{BF} = (1 - (1 - \frac{1}{L})^{n \times k})^k \]
\[ = (1 - (1 - \frac{p}{16})^{n \times k})^k \]
\[ f_{L,k,n}^{BF} \leq \eta \]
Data storage using BF

(a) \[ L = \frac{16}{p} \leq S \]

(b) \[ x = \frac{1}{p} \leq n_{max} \leq \frac{\ln(\frac{1}{1-\sqrt[k]{\eta}})}{k \ln(\frac{1}{1-p/16})} \]

Proposition 3: Given \( \eta, k \) and packet size \( S \), a proper \( p \) can be found such that requirements (a) and (b) can be satisfied. For example, if \( k=5, \ \eta=2\%, \ S=320 \), then \( p=5\% \), \( n_{max}=39 \approx 2 \cdot x > x = \frac{1}{p} = 20 \).

Rotation of The Collectors

- If the sink receives \( y \) data packets from \( y \) collectors during a round, then there are \( y \) collectors which turn their roles to COMMON.
- To maintain the data delivery ratio, \( y \) collectors should be randomly re-elected from rest common nodes. Before the sink broadcasts the VP message, it adds \( y \) into VP. When a common node \( w \) receives VP, \( w \) turns into COLLECTOR with a probability:

\[ g = \frac{y}{N-N \cdot p + y} + \frac{e_{w}}{N \cdot e} \]

Proposition 4: Let random variable \( Y \) denote the number of the re-elected collectors, then \( E(Y) = y \).
Probability analysis

Lemma 2: If $N \geq \frac{3}{p \cdot \epsilon^2} \ln(2/\delta)$, then $N \cdot p$ provides $(\epsilon, \delta)$-approximation for the number of the collectors. That is, for any $0 < \epsilon, \delta < 1$, $\Pr(|X - N \cdot p| \geq \epsilon N \cdot p) \leq \delta$.

- $N \cdot p$ is the number of the collectors’ estimation.

Probability analysis

Proposition 5: If $N \geq \frac{3}{(1-p) \cdot (1-c) \cdot \epsilon^2} \ln(2/\delta)$, then $N \cdot (1-p) \cdot (1-c)$ gives $(\epsilon, \delta)$-approximation for $X_t$, the number of the common nodes which stay in COL state during time slot $t$. That is, for any $0 < \epsilon, \delta < 1$, $\Pr(|X_t - N \cdot (1-p) \cdot (1-c)| \geq \epsilon N \cdot (1-p) \cdot (1-c)) \leq \delta$.

- one time slot, there are $N \cdot (1-p)(1-c)$ data on common nodes that can be collected, i.e., not all $N$ data items can be collected in one time slot.

$$b = \frac{\pi r^2}{Z^2} \quad \Pr(X_{w,t} = 0) = (1 - b)^{(N \cdot p)} = c$$
Probability analysis

Theorem 7: For any common node $w$, the probability that its data can be collected by collectors during $\frac{1}{1-c}$ time slots approaches to 1.

$$b = \frac{\pi r^2}{Z^2} \quad \Pr(X_{w,t} = 0) = (1 - b)^{(N \cdot p)^c}$$

Probability analysis

Theorem 9: Suppose there are $k$ hops from a collector $u$ to the sink $S$ along a path $P$ from $u$ to $S$, where $P = \{u, u_1, u_2, \ldots, u_{k-1}, S\}$. $u_1, u_2, \ldots, u_{k-1}$ are intermediate relay nodes. The probability of a successful transmission along $P$ is close to 1 if $N$ is large enough.
Simulation Evaluation

- Experiment Setup
  - NS2 with C++.
  - Mobile nodes and a sink are deployed in an area of 1200m × 1200m.
  - Random waypoint mobility model.
  - Transmission range is 100m.
  - The bloom filter with 5 random hash functions on each sensor is 320 bits. The tolerant false positive probability is 5%.

Simulation Evaluation

- Comparative state of arts protocols
  - Sidewinder [14],
  - DFT-MSN [18]
  - SCAR [19]

- Metrics
  - Data delivery ratio
  - Messages overhead
  - Time delay
Simulation Evaluation

Fig. 5. Delivery ratio vs the maximum velocity of sensor nodes.

Fig. 6. Delivery ratio vs the number of sensor nodes.

Fig. 7. Total messages vs the number of sensor nodes.
Conclusion

- The paper proposes SMITE, a novel data collection protocol for mobile wireless sensor networks based Random waypoint mobility.
- The probability analysis shows that
  - all data from the common nodes can be transmitted to the collectors with a high probability, and
  - gathered data on the collectors can also be forwarded to the mobile sink with a high probability.
- The collectors use bloom filters to compress received data.
- The simulation results on NS2 validate the theoretical results and show that SMITE outperforms the state-of-the-art schemes, DFT-MSN, SCAR and Sidewinder, on the aspects of packet delivery ratio, transmission overhead, and time delay.
Thank you so much

Q&A