Abstract—Mobile Wireless Sensor Networks (WSN) provide sensing coverage of huge areas by deploying mobile entities to collect sensing data from the static sensors and aggregate to the access points. Comparing with static WSN, such network can provide low cost, high throughput, large scale sensing coverage. However, data collection in mobile WSN strongly relies on the movement of the mobiles. Sensing data may not be collected effectively due to the limitation of the mobiles. In this paper, we introduce an energy efficient Collaborative Communication (CC) in WSN to ensure effective data collection in mobile WSN. Experiments on real taxi trajectories of 27,000 taxis in the Beijing city show that energy efficient CC decrease the energy consumption of the urban sensing network in Beijing comparing with individual communication, with an ensurance of data collection ratio larger than 80%.

Index Terms—Energy Efficient Collaborative Communication, effective data collection probability, mobile Wireless Sensor Networks

I. INTRODUCTION

The mobile Wireless Sensor Network (WSN) owns its name by introducing mobile entities into WSN to collect data from sensors distributed in the field. In recent years, mobile Wireless Sensor Network has been widely studied to sense and monitoring the physical environment, such as tracking animal migrations in remote-areas [1], measuring weather conditions in national parks [2], habitat monitoring on remote islands [3], city traffic monitoring [4] and etc. Comparing with the static WSN, mobile WSN exhibits better energy efficiency, improved coverage, and shorter data latency [5].

However, data collection in mobile WSN strongly relies on the mobility patterns of the mobile nodes. Data from the sensors may not be collected effectively due to limited number, poor coverage, and even heterogeneous distribution of the mobiles. Nevertheless, in most WSN applications, data from a certain percentage of sensors should be collected within a reasonable time constraint to reconstruct the full view of the sensing field. How can we ensure effective data collection given the limitations of the mobiles?

Different approaches have been proposed to address this problem of data collection in the mobile WSN. Intuitively, one can increase the transmission power of sensors to reach the mobile nodes. However, such approach will certainly increase energy consumption and shorten node lifetime. Mobility control approaches have also been proposed to direct the mobiles to visit specified areas [6]. However, such approaches are not practical in scenarios where uncooperative entities are acting as mobiles. Other approaches suggest adjacent sensors organizing a multi-hop ad-hoc network to relay data to sensors that can communicate with the mobiles [4]. However, such approaches will result in longer delay for data collection. In the meantime, all the approaches mentioned above are focusing on simply improving data collection in certain areas where sensors can hardly be visited by the mobiles, without ensurance of effective data collection in the whole network.

Collaborative Communication (CC) [7] [8] is an effective physical layer approach to extend the transmission range and increase energy efficiency. According to CC, a group of collaborative nodes (C-nodes) participate to transmit or receive a common signal when they communicate with an isolated node far from them. The key point in CC is to modify the carrier phase of the C-nodes so that multiple signals are synchronized [9]. For instance, in the transmit mode, multiple signals are received by the isolated node synchronously and combined constructively to increase the signal quality, or equivalently extend transmission range. CC is a feasible alternative to control the network topology in wireless networks with resource-restricted nodes, since the performance of CC mainly depends on the number of C-nodes rather than on the capabilities of each individual node. In other words, increasing in the number of C-nodes can somehow compensate the weak capabilities of individual nodes.

In this paper, an energy efficient CC scheme for mobile WSN is proposed to ensure effective data collection in the whole network. Such scheme can be utilized in large scale sensing networks and intelligent remote meter reading systems, in which effective data collection ratio is required to guarantee the functionality of the network. Experiments on Beijing taxi trajectories is further deployed to verify the energy efficiency and data collection performance of the scheme.

The remainder of this paper is organized as follows. In Section II, we introduce the network model and the energy efficient CC scheme. The expressions of OTR for different mobility models are discussed in Section III. Experiments on real taxi trajectories are presented and analyzed in Section IV. And the conclusions are drawn in Section V.

II. NETWORK MODEL

Suppose there are $N$ clusters of static sensors randomly distributed in a sensing field, modeled as a unit disk of radius 1. Each cluster includes $L$ sensors, the maximum distance between any pair of nodes in the same cluster is denoted by $d_0$. 

\begin{align*}
N & \text{clusters of static sensors} \\
L & \text{number of sensors in each cluster} \\
d_0 & \text{maximum distance between nodes in the same cluster}
\end{align*}
These sensors can either transmit their data to a closeby mobile by themselves, indicated as the Individual Communication (IC) mode, or collaborate with other sensors in the same cluster, indicated as the Collaborative Communication (CC) mode, to accomplish the data transfer task. The sensors choose to transmit in either mode to minimize the energy consumption of the network. 

$M$ mobile nodes are moving in the field. The communication range between sensors and mobile nodes is denoted by $r$, which is achieved either by the IC mode or by the CC mode. Assume $d_0 \ll r$, so that the range can be calculated from the geographic center of the clusters in the CC mode. A sensor is considered collected when its data is successfully transmitted to a mobile node within the communication range. Note that in a mobile WSN, the mobile nodes can be in the communication range of the C-nodes for only a short time. Therefore, we assume the number of C-nodes $L$ in each cluster is predetermined and fixed, so that synchronization and data aggregation can be done in advance to ensure timely data transmission to the mobiles.

The network is considered functional if the probability of all the sensors being collected by the mobile nodes at least once within the time constraint $T_{MAX}$ is larger than $C$.

$$P = P \{ \bigcap_{i=1}^{NL} A_i \} > C,$$  

where $A_i$ is the event that sensor $i$ is collected at least once by the mobiles within $T_{MAX}$. The assumption is due to the fact that in most WSN applications, only one sample from a sensor in a certain period of time is needed to reconstruct the sensing field. In the meantime, data from a certain percentage of sensors should be sufficient to get the full view of the field, therefore we have the probability constraint $C$. Note that the following optimization is conducted on the condition that $P > C$ rather than $P \{ A_i \} > C$, since the probability constraint on $P \{ A_i \}$ may not guarantee an effective data collection probability larger than $C$.

The energy consumption of each sensor for transmitting $T_d$ duration of data stream can be expressed as follows,

$$E = \min \{ E_{IC}, E_{CC} \},$$

where $E_{IC}$ and $E_{CC}$ indicate the energy consumption of each sensor in the IC and CC mode, respectively.

The energy consumption for the IC mode is

$$E_{IC} = T_d P_{tIC}^c,$$  

where $P_{tIC}^c$ is the power consumption for data transmission from the sensors to the mobile nodes in the IC mode.

We utilize the CC scheme introduced in [10] for the CC mode, in which optimal number of C-nodes are selected in the cluster to perform the collaborative communication so as to minimize the energy consumption. Therefore the energy consumption in the CC mode is

$$E_{CC} = L T_d P_{rCC}^c + (2L + 1) T_m P_{tc} + L T_d P_{tc},$$

where $T_m$ is the duration of local communication message to initiate CC. $P_{tc}$ is the local transmission power consumption amongst the C-nodes, and is a function of $d_0$. $P_{tCC}^c$ is the power consumption of each C-node for CC from the C-nodes to the mobile nodes in the CC mode.

$P_{tIC}^c$ can be calculated as follows [11],

$$P_{tIC}^c = \frac{\xi}{\eta G_t G_r} \left( \frac{4\pi}{\lambda} r^2 k + P_c \right),$$

where $\xi$ is the Peak to Average Ratio (PAR), $\eta$ is the drain efficiency of the RF power amplifier, $P_{tCC}^c$ is the effective receiving power in the IC mode. $G_t$ and $G_r$ are the antenna gains for the transmitter and receiver, respectively. $\lambda$ is the wavelength of the transmitted signal. $k$ is the path loss exponent. $P_c$ is the total power consumption of the D/A converter, the mixer and the active filter. $P_{tCC}^c$ and $P_{tc}$ can also be derived similarly by plugging the corresponding receiver powers and distances into (5).

There exists a critical distance $r_0$ in CC. When $r \geq r_0$, $E_{CC} < E_{IC}^c$, and it is energy efficient to utilize CC, while when $r < r_0$, $E_{CC} > E_{IC}^c$, indicating that the energy saved through collaboration can not justify the energy used for local communication and data aggregation, and it is not energy efficient to use CC anymore [12]. Therefore,

$$E = \begin{cases} E_{IC} = T_d \left\{ \frac{\xi}{\eta G_t G_r} \left( \frac{4\pi}{\lambda} r^2 k + P_c \right) \right\}, & r < r_0 \\
E_{CC} = L T_d \left\{ \frac{\xi}{\eta G_t G_r} \left( \frac{4\pi}{\lambda} r^2 k + P_c \right) \right\} + (2L + 1) T_m P_{tc} + L T_d P_{tc}, & r \geq r_0 \end{cases},$$

where $P_{tIC}^c$ and $P_{rCC}^c$ are the efficient power for receivers in the IC and CC mode, respectively.

(6) shows that the energy consumption is a monotonic increasing function of the communication range $r$. Therefore, to minimize $E$, we only need to minimize $r$.

$$\min E \iff \min r \quad \text{subject to } P > C$$

**Definition 1 (Optimal Transmission Range (OTR))** The Optimal Transmission Range is defined as the minimum communication range that satisfies $P > C$, i.e. $OTR = \min r$. Given OTR for a certain scenario, we can always derive the optimal energy consumption through (6). Therefore, in the following section, we will further discuss the OTR under different mobility models.

### III. Optimal Transmission Range

The OTR for Random Mobile Model (RM) and Random Waypoint Model (RWP) [13] are listed in Table. I. Detail process of derivation can be found in the appendix. These expressions can be utilized to estimate OTR in realistic scenarios, in which the distribution of the location of the mobiles is similar to that of either model.

In the RM model, mobile nodes perform random walk on the unit disk, resulting in a uniform location distribution at each time slot, while in the RWP model, mobile nodes choose their waypoints with a uniform distribution over the
unit disk, and the nodes then move to the waypoint with a velocity drew from a uniform distribution in \([V_{\text{min}}, V_{\text{max}}]\), independent of the node location. \(f(l)\) is the pdf of the location of mobile nodes following the RWP model, in which \(l\) indicates the distance of the mobile from the center of the unit disk. Normally, the transmission pattern of the CC mode is directional with transmission angle of \(\theta \in [0, 2\pi]\), as an advantage of collaborative beamforming, while that of the IC mode can be considered omnidirectional (\(\theta = 2\pi\)), as a result of the simplicity of sensor nodes.

Fig. 1 shows the OTR results of the simulation experiments of RM Model and RWP Model, and compares them with the theoretical results. Each simulation curve is the result of \(10^8\) simulation experiments of mobile nodes moving on the unit disk following the two models respectively. The OTR is then calculated according to the definition as the transmission range that results in a frequency of the clusters being visited by the mobiles larger than \(C = 0.8\).

The Mean Square Error (MSE) of the theoretical results comparing with the simulation results is

\[
E\{(OTR_{\text{theory}} - OTR_{\text{simulation}})^2\} = 0.0061
\]

for RM Model, and

\[
E\{(OTR_{\text{theory}} - OTR_{\text{simulation}})^2\} = 4.5302 \times 10^{-4}
\]

for RWP Model, respectively, which in comparison with the values of OTR are negligible. Such results verify that the theoretical expressions are comparatively accurate to estimate the OTR in both models. Similar verifications can also be derived for multi cluster cases.

Fig. 2 shows the OTR in terms of time constraint for \(N = 5\) and \(N = 1\). In \(N = 5\) case, sensor clusters in the RWP Model simulations are located on a circle with its center on the original point of the unit disk. From these results we can see that in both mobility models, the OTR decreases as the time constraint increases. This is reasonable since if there are more time for the mobile nodes to travel, they will have more chance to visit the sensors even if the transmission range is small.

The OTR of RWP Model is directly related to the location of the sensors. From Fig. 2 we can see that the OTR of RWP model when the sensor clusters are located at \(l = 0.5\) is smaller than the OTR of RM Model in both cases; while the OTR of RWP model when the sensor clusters are located at \(l = 0.8\) is larger than that of the RM Model. That is because, according to the pdf of the RWP model, mobile nodes are more likely to be distributed near the center of the unit disk. Therefore, if the sensors are located near the center, the OTR of the sensors can be smaller to satisfy the data collection constraint.

### IV. Experiments on Beijing Taxi Trajectories

The Beijing taxi trajectory database contains 30-day GPS records of 27,848 taxis traveling within 39.759°N to 40.023°N latitude and 116.209°E to 116.544°E longitude, which is approximately the area within the fifth-ring road in Beijing. In the experiments, sensors are uniformly distributed in Beijing, and grouped into clusters. Taxis are acting as mobiles to collect data from the static road-side sensors. The following experiments are conducted using the real traces of taxis on May 1, 2009, after invalid traces being removed.
Fig. 3. Heatmap of Data Collection without Collaboration in Beijing

Fig. 3 is the heatmap of data collection without sensor collaboration. The color of each point indicates the number of data collected from the sensor at corresponding location in Beijing. As is shown, sensors in the city center and near the main roads are more likely to be visited and get their data collected, while “cold spots” exist in suburbs and large parks where data can hardly be collected. Statistics show data collection percentage is lower than 50% when the transmission range is 100m for each sensor and nearly 5,000 taxis are utilized to collect data from 8am to 10am on May.1st, 2009.

We then utilize energy efficient CC in the network to achieve data collection percentage of 80%. We assume $N_0 = -171dBm$, $G_t = G_r = 2dBi$, carrier frequency $f = 2.4GHz$, bandwidth $B = 10K$, $\xi = 2$ for QPSK modulation, $\eta = 0.35$, $T_d = 0.5s$, $T_m = 0.1ms$, $d_0 = 50m$, $k = 3.5$ for local transmission, and $k = 2$ for long-haul transmission.

The OTR and energy consumption of the network in terms of taxi numbers are shown in Fig. 4. We use the RWP model to estimate the OTR in such network, and compare with the OTR drew from the taxi trajectory dataset. Due to the similarity of the location distribution of the mobiles in RWP model with that of the taxis in the city, the estimated OTR drew from RWP model is quite close to the OTR derived from the real traces.

The energy consumption of the sensors is proportional to the transmission range. The energy consumption curves show that utilizing CC for data collection in such urban scale WSN can sufficiently decrease energy consumption with the assurance of effective data collection comparing with individual communication. In the meantime, increasing $L$ does not necessarily decrease energy consumption, since more C-nodes will bring extra energy consumption for local communication and data aggregation. Given OTR, future work can be done to further optimize energy consumption by carefully choosing $L$.

Fig. 5 is the heatmap of data collection using the same trajectories in Fig. 3, when energy efficient CC is utilized. As is shown, data collection is increasing effectively and data from the main part of the city is collected.

V. CONTRIBUTIONS

In this paper, an energy efficient collaborative communication scheme is proposed to minimize energy consumption in mobile WSN with respect to effective data collection. We further derive the optimal transmission range to achieve the minimized energy consumption in the Random Mobile Model and Random Waypoint Model, which can be utilized to estimate OTR in real deployments. Experiments on Beijing Taxi Trajectory dataset is conducted to verify the energy efficiency of utilizing collaborative communication in large scale WSN with assurance of data collection ratio larger than 80%.
A. OTR for Random Mobile Model

1) \(N=1\): Suppose there is only one cluster of sensors deployed randomly on the unit disk, of which the transmission range of the C-nodes is \(\frac{\theta r^2}{2\pi} = \frac{\theta r^2}{2\pi}\). So the probability that within time constraint \(T_{MAX}\), the sensors will communicate with at least one mobile node is as follows.

\[
P(N = 1) = 1 - (1 - \frac{\theta r^2}{2\pi})^{MT_{MAX}},
\]

According to the definition of OTR, \(OTR = \min_{P>C} r, 0 \leq r \leq R\).

\[
P(N = 1) = 1 - (1 - \frac{\theta r^2}{2\pi})^{MT_{MAX}} > C
\]

\[
\Rightarrow r > \sqrt{\frac{2\pi}{\theta} (1 - \frac{MT_{MAX}}{1 - C})}.
\]

Therefore, the optimal transmission range in this case is

\[
OTR = \sqrt{\frac{2\pi}{\theta} (1 - \frac{MT_{MAX}}{1 - C})}.
\]

2) \(N > 1\): Suppose there are \(N\) clusters of sensors randomly deployed on the unit disk \((N > 1)\). Define the event that these \(N\) clusters can all be visited at least once by the mobile nodes within deadline as \(A\). The complementary set of \(A\) is the union set of \(A_i\),

\[
P(\overline{A}) = P(\bigcup_{i=1}^{N} A_i),
\]

where \(A_i\) is the event that cluster \(i\) can be visited at least once within deadline. Therefore, the probability of the union set can be derived as [14]

\[
P(\overline{A}) = P(\bigcup_{i=1}^{N} A_i) = \sum_{n=1}^{N} (-1)^{n-1} C_N^n a_n,
\]

where \(a_n\) is the probability of the intersection of any \(n\) events \(A_i\),

\[
a_n = P(\bigcap_{i \in I} A_i), I \subset 1, 2, \cdots, N, |I| = n,
\]

\[
a_n = (1 - n \frac{\theta r^2}{2\pi})^{MT_{MAX}}.
\]

Therefore,

\[
P(\overline{A}) = P(\bigcup_{i=1}^{N} A_i) = \sum_{n=1}^{N} (-1)^{n-1} C_N^n a_n
\]

\[
= \sum_{n=1}^{N} (-1)^{n-1} C_N^n (1 - n \frac{\theta r^2}{2\pi})^{MT_{MAX}}.
\]

So, the probability that the \(N\) clusters can be visited at least once within \(T_{MAX}\) is

\[
P(N > 1) = P(A) = 1 - P(\overline{A})
\]

\[
= 1 - \sum_{n=1}^{N} (-1)^{n-1} C_N^n (1 - n \frac{\theta r^2}{2\pi})^{MT_{MAX}}.
\]

And \(OTR = \min_{P(N > 1) > C} r, 0 \leq r \leq R\). Given \(N, M, T_{MAX}\) and \(C\), we can always find the optimal \(r\).

B. OTR for Random Waypoint Model

The pdf of the mobile node location following the RWP model on a unit disk is as follows [15].

\[
f(l) = \frac{45(1 - l^2)}{64\pi} \int_{0}^{\pi} \sqrt{1 - l^2 \cos^2 \phi} d\phi,
\]

where \(l\) is the distance of the location to the original point. The pdf stated above can be approximated with the polynomial of the form

\[
P_1(l) = \frac{2}{\pi} (1 - l^2),
\]

with the MSE of \(6.5 \times 10^{-4}\) [16]. Fig. 6 shows the pdf of the location of the mobile nodes following the RWP model. As we can see, the mobile nodes are more likely to be in the center of the disk than on the borders. Such characteristic can be used to approximate the distributions of cars in the cities, since cars are also more likely to be in the city center than in the suburbs.

When talking about the arrival process of the mobile nodes of a coverage area, we can even make a local Poisson assumption and simply assume that, for a single mobile node, the arrival process of this node within the area \(B_r(l)\) results from a homogeneous Poisson point process with mean equal to \(f(l)B_r(l)\) [17], in which \(|B_r(l)|\) indicates the area of \(B_r(l)\). Since the movement of each node is independent of each other, the superposition of \(M\) mobile nodes also follows a Poisson distribution with mean equal to \(M|B_r(l)|f(l)\).
(22)

The OTR of this case is then derived as the minimal transmission range that satisfies $P(N = 1) > C$.

$$\text{OTR} = \sqrt{-2 \ln(1 - C) \over M_{T_{\text{MAX}}} \theta_f(l)}.$$  \eqref{eq:23}

2) $N > 1$: When there are $N$ clusters located in the field, we can consider the arrival rate of the coverage areas of these $N$ clusters independent of each other. Therefore, the probability that these $N$ clusters can be visited at least once by the mobile nodes within deadline is

$$P(N > 1) = \prod_{i=1}^{N} P_i = \prod_{i=1}^{N} \left(1 - e^{-M_{T_{\text{MAX}} \theta_f(l)}}\right),$$  \eqref{eq:24}

in which $P_i = 1 - e^{-M_{T_{\text{MAX}} \theta_f(l)}}$ is the probability that cluster $i$ ($i = 1, 2, \ldots, N$) will be visited by the mobile nodes at least once within $T_{\text{MAX}}$. The OTR can then be derived by calculating the minimum $r$ that satisfies $P(N > 1) > C$.

In particular, if all the sensor groups are located on a circle with the center on the original point, all the sensors share the same $f(l_i)$. In this case, $P(N > 1)$ can be simplified as

$$P(N > 1) = \prod_{i=1}^{N} P_i = \left(1 - e^{-M_{T_{\text{MAX}} \theta_f(l)}}\right)^N,$$  \eqref{eq:25}

and

$$\text{OTR} = \sqrt{-2 \ln(1 - \sqrt[N]{C}) \over M_{T_{\text{MAX}}} \theta_f(l)}.$$  \eqref{eq:26}

ACKNOWLEDGEMENT

This work is partially sponsored by the Karlsruhe House of Young Scientists (KHYS).

REFERENCES


with a vehicular sensor network,” IEEE Wireless Communications, October 2006.