The Base Station Placement for Delay-constrained Information Coverage in Mobile Wireless Networks

Gabriel Y. Keung, Qian Zhang, and Bo Li
The Hong Kong University of Science and Technology

Abstract—In this paper, we study the delay-constrained information coverage problem in mobile wireless networks. Motivated by real application needs, our formulation takes advantage of the node mobility for information collection, which takes place when a node moves into the proximity of stationary base stations. To the best of our knowledge, we present the first formulation for the delay-constrained information coverage problem, which targets at optimal base stations placement with the objective of maximizing information collection within a constrained time. We prove that this problem is NP-hard and we develop theoretical analysis to derive its upper and lower performance bounds. We then develop approximation techniques and use simulations to verify their effectiveness.

Keywords—Mobile Wireless Network, Mobility, Information Coverage, Base Station Placement.

I. INTRODUCTION

Our study is motivated by the data message collection over a region by using mobile nodes. In such application, a device (e.g., Pocket PC) having communication ability is attached to moving objects for message delivery. However, due to the privacy requirement, multi-hop transmissions across different mobile nodes for sensing information relying from moving objects to base stations should not be used in this scenario. Therefore, by taking advantage of the uncontrolled node mobility, a basic strategy would be to only allow information delivery when sensors are within the transmission (coverage) region of a base station during their movement.

In such applications, the completeness and timeliness of information collection become two important criteria. This implies that the coverage from base stations for information collection needs to be maximized within a constrained time. The basic problem can be formulated as a mobile wireless network consisting of a potentially large number of uncontrolled mobile nodes and a single or multiple stationary base stations for information collection. This information collection has to take into account the time-sensitive nature of the data message. Under this scenario, the information coverage problem can be defined as to maximize the percentage of sensing information from covered mobile nodes by proper base station placement. Given the time-sensitive nature of the information, a delay-constrained information coverage problem can be then defined as to maximize the percentage of information from covered nodes by proper base station placement during a time period.

In this paper, we formally describe the delay-constrained information coverage problem in mobile wireless networks with the random walk mobility models. The solution for this problem critically depends on the initial placement of base stations and the mobility pattern of nodes. The optimal solution for this problem turns out to be difficult due to the proved NP-hard nature of the formulation and the infinite searching space of base station positions. We first restrict the base station position from an infinite space to a finite space and apply the Simulated Annealing (SA) based algorithm to solve the combinatorial optimization problem. We then develop theoretical analysis to derive the upper and lower bounds on the performance. Finally, we use simulations to evaluate and verify the effectiveness of the proposed algorithm.

The rest of this paper is organized as follows. Sec. II reviews the related works. Sec. III describes the delay-constrained information coverage problem and proposes an approximation algorithm based on restriction from an infinite searching space and the Simulated Annealing. Sec. IV examines the performance of the proposed algorithm with simulations, and finally Sec. V concludes the paper.

II. RELATED WORKS

Recently, information collection in mobile wireless networks has received increasing attentions, especially in mobile sensor networks. Several schemes have been introduced, which can be broadly divided into two categories: epidemical-based and probabilistic delivery approaches.

In [1], an epidemical-based approach was proposed. The basic idea is to randomly spread the information over the entire network so that eventually a sink could be reached. Small et al. described an interesting application of epidemic routing protocols using whales as carriers for cost-effective information collection [2]. In general, this approach can achieve excellent information delivery ratio with the assistance of sufficiently large buffers, but the price to pay is the potentially high cost in terms of the communications and energy consumptions.

In [3], a probabilistic delivery approach was presented, which enabled asynchronous communications among intermittently connected clouds of sensor nodes. Wang et al. introduced a probabilistic delivery approach [4], which examined how the data replication in sensor networks can be constrained using a fault tolerance value associated to each data message. This scheme could result in significantly amount of data replication and communication overhead associated with replication node selections.

*The research was support in part by grants from RGC under the contracts 615608, and 616207, by a grant from NSFC/RGC under the contract N_HKUST603/07. This work was also supported in part by National Natural Science Foundation of China with grant no. as 60933012.
Our approach in this paper is completely different from all prior schemes. The direct information collection incurs the minimum communication overhead and fulfills privacy requirement by eliminating the need for forwarding information. The challenge, however, is to maximize such direct communications for information collection between mobile nodes and base station(s) within a constrained time.

III. Problem Description and Formulation

We formally describe the delay-constrained information coverage problem and the formulation of the optimization.

A. Problem Description

We consider a mobile wireless network consisting of \( M \) mobile nodes. Nodes follow a given mobility model within a 2-D geographic area \( A \) to store and deliver some data messages. There exist a single or up to \( N \) stationary base stations for collecting information. Let the coverage range of a base station to be \( R \), and its coverage region to be specified by \( S_h \), \( (S_h \in A, 1 \leq h \leq N) \). This implies that a base station is capable of collecting information from mobile nodes located within the disk of radius \( R \) centered at the base station. Hence, once a mobile node entering the coverage region of a base station, we say that the mobile node is covered. However, it is not necessary meaning that the base stations can be deployed on any positions inside \( A \), we can further added deployment constraints to limited the possible region for base station deployment based on the physical situation.

Without losing the generality, we assume the probability distribution of mobile nodes’ initial geographical positions are known, referred as the node position distribution. Based on this distribution, we can calculate which area has higher probability to find a mobile node (or a group of mobile nodes). Specifically, we assume that, at the initial time instant \( t_0 \), the node position distribution \( D(A, t_0) \) within the area \( A \) follows a certain pattern according to different application scenarios. We will further elaborate and specify in this section.

We can now formally define the delay-constrained information coverage problem as follows:

**Definition 1.** We first define an probability that a mobile node is covered by base stations at a particular time \( t \), i.e., the probability that the mobile node is within the coverage region of any base station at a time \( t \), referred as the coverage probability. We then define the information coverage of mobile nodes over a period \( T \), denoted by \( \text{Coverage}(S_h, T) \), as a random variable (an integer from 0 to \( M \cdot T \)) representing the number of times for mobile nodes entering the coverage region of any base station \( h \) during the period \( T \). The average coverage probability for the network can then be calculated by the information coverage of mobile nodes over a period \( T \) divided by the number of nodes \( M \) and the period \( T \).

The coverage formulation captures the information collection within a time interval, a higher coverage value implies better collection performance; alternatively, in order to collect certain amount of information, the shorter time used to meet the requirement, the higher performance we could obtain.

We assume that each mobile node performs the 2-D random walk movement, which is one of the most common and widely used mobility models [5]. With this mobility model, each mobile node travels from its current location to a new location by randomly choosing a direction \( \theta \in [0, 2\pi) \) and a speed \( v \in [v_{\text{min}}, v_{\text{max}}] \), respectively, in each discrete time interval \( \Delta t \). We further make one simplification that all mobile nodes move at a constant speed \( v \), so that the distance travelled in each discrete time interval can be denoted by \( r_v = v \Delta t \). Under the random walk mobility, the probability \( P(r, \theta, t) \) at particular time \( t \) can be calculated from the initial node position under the polar coordinate, which is described in [6]. We can further approximate the probability \( P(r, t) \) with radial symmetric \( \theta \) by the Gamma distribution:

\[
P(r, t) = \frac{r^{k-1} \beta^k \exp(-r \beta)}{\Gamma(k)},
\]

where \( \Gamma(k) \) is the Gamma function of \( k \), as well as \( k \) and \( \beta \) are time \( t \) related shape and scale parameters for the Gamma distribution, respectively.

B. Restraint to Finite Searching Space

Although the area size \( |A| \) is bounded, the searching space of base station placement is infinite. There exists an infinite number of points in the searching area and a native approach such as breaking up the area into uniform grid subareas cannot provide any properties that can be used to calculate the information coverage. To obtain a finite searching space, we use the concentric ring areas \( C_k(k \geq 1) \) of a mobile node illustrated in Fig. 1. Inside the same concentric ring area, the probability to locate a mobile node is the same, which is the average value of \( P(r, \theta, t) \) over the area. Concentric rings from different mobile nodes intersect with each other and divide the area \( A \) into small subareas. With this approach, we can take the advantage of the radial symmetric \( \theta \) from the node mobility, and directly apply the approximations of the probability \( P(r, \theta, t) \) for the random walk and the Gauss-Markov mobility models into the subareas.
The remaining \((M - M_{0,h})\) mobile nodes, which are initially expected to be outside the coverage region \((S_h)\) of a base station, can be classified to subset of \(\bigcup_j \cdots \bigcup_i M_{i,h}\). \(M_{i,h}\) denotes the set of mobile nodes that initially have the same Euclidean distance \((i \cdot r_v : \left[ \frac{R}{r_v} \right] \leq i \leq \left[ \frac{T}{r_v} + 1 \right] \) from the position of the base station \(h\), and \((M - M_{0,h})\) is equal to \(\sum_i |M_{i,h}|\).

The normalized probability function of the remaining expected \((M - M_{0,h})\) mobile nodes then becomes \(\sum_{k=1}^N P(C_k, t_j) = 1\), where those concentric ring areas \(C_k (k \geq 1)\) are denoted as the area of possible locations of a node as time goes by, as illustrated in Fig. 1. Here we can directly apply the approximation of the random walk mobility from (1) to the probability \(P(C_k, t_j)\). Note that at time \(t_j = t_0 + j \Delta t\) the farthest Euclidean distance travelled by a node cannot exceed \(C_j\) under the random walk model. Thus, the information coverage of a mobile node, which located at the Euclidean distance \(i \cdot r_v\) from the base station, with the time interval \([t_0, t_0 + T]\) can be expressed as follows:

\[
Q_h(i, T) = \sum_{j=1}^{T/\Delta t} P(C_i, t_j) \cdot L(C_i \cap S_h) + \ldots + \sum_{j=T/\Delta t}^{T/\Delta t} P(C_T, t_j) \cdot L(C_T \cap S_h),
\]

where the probability term \(L(C_k \cap S_h) \geq 0\) is related to the cross section area between the possible locations of a mobile node and the coverage region \(S_h\), as shown in Fig. 1.

The total information coverage can be calculated by directly sum-up the expectation of information coverage from every base station. If the base stations do not overlap with each other, which means \(S_j \cap S_k = \emptyset\) for \(k \neq j \in \aleph\), then the overall expectation of information coverage is following:

\[
\text{Max} \quad \sum_{h=1}^N E[\text{Coverage}(S_h, T)]
\]

s.t.

\[
E[\text{Coverage}(S_h, T)] = \sum_i |M_{i,h}| \cdot Q_h(i, T) + M_{0,h}
\]

\[
\left( \left[ \frac{R}{r_v} \right] \leq i \leq \left[ \frac{T}{r_v} + 1 \right] \right)
\]

\[
Q_h(i, T) = \sum_{j=1}^{T/\Delta t} P(C_i, t_j) \cdot L(C_i \cap S_h) + \ldots + \sum_{j=T/\Delta t}^{T/\Delta t} P(C_T, t_j) \cdot L(C_T \cap S_h)
\]

\[
\sum_{k=1}^j P(C_k, t_j) = 1 \quad (t_0 \leq t_j \leq t_0 + T)
\]

\[
S_h \in A \quad (h \in \aleph, T \geq 0)
\]

The approximation reduces the searching space of base station placement to a finite number of subareas, in which we can quantify each subarea in terms of the different probabilities \(P(C_k, t_j)\). We make a simplification that the probability \(P(C_k, t_j)\) is approximated by the average value of \(P(r, \theta, t)\) over the same concentric ring area \(C_k\). The accuracy of the probability \(P(C_k, t_j)\) is related to the distance that a mobile node travels in each discrete time interval \(\Delta t\). To improve the calculation accuracy, one can shorten the unit of discrete time intervals, which results the increment of number of subareas.

**Lemma 1**: The delay-constrained information coverage problem with restricted search space under the random walk mobility model is NP-hard.

**Proof**: This formulation can be considered as a restricted case in that nodes are all static with \(k\) number of base stations; then the information coverage problem is reduced to the Geometric \(k\)-Clustering problem [7], which is proved to be NP-hard. Therefore, we conclude that the original formulation is also NP-hard.

**C. Optimal Base Station Placement Algorithm**

Before we introduce the Simulated Annealing (SA) algorithm for base station placement, we first use a simple exhaustive algorithm to solve the delay-constrained information coverage problem. This allows us to obtain some understanding on the basic characteristics of the optimal solution. Fig. 2 presents the optimal solution for a single base station, obtained by exhaustive search for a hundred mobile nodes under the random walk model. The exhaustive search algorithm scans through the uniform grid subareas in \(A\) to find out a position that results in the maximum total probability (black colored regions) inside the coverage region of a base station. Fig. 2 also shows that the 2-D probability distribution consists of many local maxima and disperses through the area (black colored regions inside white and gray regions); this implies that a simple iterative improvement algorithm may terminate in first found local maxima, which does not necessarily lead to the optimal solution in this case.

We adopt the SA technique which is categorized as an iterative improvement algorithm [8], commonly used for combinatorial optimization problems. The SA algorithm includes an acceptance probability, which can prevent it from terminating at local minima/maxima. The algorithm works as following:

- Randomly place the \(N\) number of base stations inside the area \(A\). Since \(A\) is divided into several subareas after applied the restriction in searching space, for each subarea only one base station is placed. The rationale behind this is to prevent overlap of the coverage regions \(S_j \cap S_k = \emptyset\) for \(k \neq j \in \aleph\), which will diminish the overall information coverage.
- With the random placement of base station(s), the first step is to improve the information coverage from the initial placement. Each base station calculates the current information coverage, and searches all unoccupied neighboring subareas for the SA algorithm.
- At each step, the SA algorithm compares the information coverage of neighboring subarea \(s'\) with the current \(s\). The information coverage difference \(\Delta\) between current position and possible candidates then can be calculated.
- A base station either will move to a position with largest information coverage improvement or move to a new
neighboring subarea with probability \( \exp(-\Delta/\text{temp}) \). \( \exp(-\Delta/\text{temp}) \) is called the acceptance probability and is related to the information coverage difference \( \Delta \) and temperature \( \text{temp} \) in the SA algorithm.

- The steps are repeated until the stop criterion is satisfied. For example, a given computation budget has been exhausted or further improvement is no longer possible.

Fig. 3 shows the optimal placement for six base stations with a hundred mobile sensors under the random walk mobility model, which indeed confirms that the coverage region of six base stations cover the highest probabilities regions.

D. The Upper and Lower Bounds on the Performance

In this subsection, we examine the upper and lower bounds on the information coverage performance.

From the previous formulations, we can calculate the probabilities \( P(r, \theta, t) \) and their intersections (subareas) \( L(S_i) \geq 0. \)

We assume the \( i \)-th highest probability density \( P(i) \) of subarea \( S(i) \) and \( f(i) \) for the fractional value \( \in (0, 1) \) in subarea \( S(i) \).

In the multiple base stations case, the optimal solution \( OPT \) is bounded by the total probability in the highest \( K' \) subareas: \( OPT \leq \sum_{j=1}^{K'} P(j) \), where \( j \in \mathbb{R} \) and the relationship between \( K' \) and \( N \) is: \( \sum_{j=1}^{K'} S(j) = N \pi R^2 \), then we have the upper bound \( (B_{max}^m) \) of the information coverage problem for \( N \) number of base stations:

\[
OPT \leq B_{max}^m = N \pi R^2 \left( \sum_{j=1}^{K'} P(j) \right),
\]

This shows the ideal case, in which are base stations covered with the highest probability subareas up to \( S(K') \).

On the other hand, a lower bound \( (B_{min}) \) can be introduced for the SA algorithm from (2), which is the minimum information coverage among all of the local maxima from the 2-D probability distribution (shown in Fig. 2). If a random base station placement is applied, \( B_{min} \) can be calculated by \( N \pi R^2 \cdot \bar{P} \).

Then the \( B_{min} \) on the performance is:

\[
B_{min} = \max \left\{ \sum_h \left[ \sum_{M_{0,h}} \int D(A-S_h,t_0) \cdot Q_h(T) \, dt \, d\theta \right] \right\}
\]

where \( Q_h(T) \) times \( D(A-S_h,t_0) \) results in the minimum of the local maxima of the information coverage within a time interval \( T \), with \( (M - M_{0,h}) \) mobile nodes, and \( \bar{P} \) is the average probability distribution over the 2-D area \( A \).

IV. NUMERICAL RESULTS AND DISCUSSIONS

In this section, we present numerical results and discussions to demonstrate the effectiveness of the proposed algorithm. We use simulations for verification and also examine the coverage performance with respect to the upper and lower bounds.

We develop a simulator that captures the essential aspects of the mobile wireless network described in Sec. III. The simulator provides the flexibility in selectively changing the configuration by setting various parameters including: (i) the area size \(|A|\); (ii) the number of mobile nodes \((M)\); (iii) different types of the node position distribution across the area \( A \) at time \( t_0 \), e.g., point distribution or/and disk distributions; (iv) the speed of mobile nodes \((v)\); (v) the coverage range of a base station \((R)\); (vi) the number \((N \geq 1)\) and positions of base station; (vii) the length of the time interval \((T)\); (viii) the mobility models of nodes, e.g., the random walk or the Gauss-Markov models. Unless otherwise specified, we use the following default settings: we define 100 mobile nodes randomly distributed in an area of size 100 \times 100 with the mobile node speed \( v = 1 \) per unit time and the coverage range \( R = 10 \) of a base station.

Starting from a specific initial distribution of mobile nodes, the simulator calculates the average coverage probability of mobile nodes, which is defined in Sec. III.A (that is, \( \sum_h Coverage(S_h,T)/(T \cdot M) \)), under a given number of base stations and their initial positions.

First, we examine the performance differences in term of the average coverage probability from the proposed approximation algorithm, the upper and lower bounds obtained from the theoretical calculations from (3-5), respectively. Fig. 4 plots the comparison of the coverage probabilities by varying the number of base stations from 1 to 27, under the random walk mobility, and the point distribution with a time interval \( T = 500 \). The average coverage probability of the approximation algorithm outperforms the theoretical lower bound and is limited by the upper bound. It shows the comparable information coverage of the SA algorithm to the upper bound when the base station number is small, as the highest probability density can be found and easily covered by SA algorithm. However, the information coverage drops considerably when the number of base stations increases; this is due to the shape limitation of the coverage region discussed in Sec. III.D.

After the theoretical calculations, we use simulations to examine the performance of the proposed algorithm. We try to compare the information coverage with the theoretical upper bound, the optimal solution (by an exhaustive maximum searching) and the random base station placement. A thousand rounds of simulation has been performed and the average coverage probabilities are plotted out. Fig. 5 shows the comparison of the coverage probabilities by varying the number of base stations from 1 to 10, under the random walk mobility, and the initial disk distribution \((d = 10)\) with a time interval \( T = 500 \). The performance of the proposed algorithm is very close to the result from the exhaustive maximum searching, while is limited by the upper bound and is far better than the random placement.

We next examine the efficiency of the SA algorithm. Fig. 6 plots the average coverage probability of three, six and nine base stations placement under the SA algorithm against iteration rounds. It is reasonable that the nine base stations placement needs more iteration rounds to be stabilized. And all placement can be stabilized within 150 iterations.

In order to demonstrate the generality of the proposed approximation algorithm, a longer time interval \((T = 2000)\), the initial disk distribution \((d = 10)\) and the Gauss-Markov...
mobility are applied for simulations. Fig. 7 illustrates the comparison of the coverage probabilities by varying the number of base stations from 1 to 10, under the random walk mobility, and the point distribution with a time interval \( T = 2000 \). The figure demonstrate the performances of the proposed approximation algorithm are very close to the results from the exhaustive maximum searching.

V. CONCLUSIONS

In this paper, we present and study the delay-constrained information coverage problem in mobile wireless networks, motivated by real applications. By taking advantage of the mobility of nodes, direct communications between a mobile node and a base station can take place when the mobile node moves within the proximity of the base station. We formulate this problem into an optimization problem with the objective of maximizing the information coverage by finding optimal placement of the base stations. We first show that the problem is NP-hard. We then restrict the base station position from an infinite space to a finite space and apply Simulated Annealing based algorithm to solve the combinatorial optimization problem. We further develop theoretical analysis to derive the upper and lower bounds on the performance.

There are several avenues for further study on this problem: (1) to consider other mobility models, e.g., the group mobility model; (2) to study the trade-off between the number of base stations and information coverage performance; (3) to consider other node position distributions, more realistic area shapes and situations (for example, with obstructions).

REFERENCES