www.s2is.org
International Journal on Smart Sensing and Intelligent Systems
ISSN 1178-5608

UTILITY BASED DATA GATHERING IN MOBILE SENSOR NETWORK

Liu Jieyan[1], Wu Lei[2], Gong Haigong [1]

[1] School of Computer Science and Engineering

University of Electronic Science and Technology of China

Chengdu, China

[2] School of Mathematical Sciences

University of Electronic Science and Technology of China

Chengdu, China

Email: Liujy@uestc.edu.cn

Submitted: Oct. 9, 2012 ***Accepted: May 10, 2013*** ***Published: June 5, 2013***

Abstract - Traditional data gathering approaches cannot be applied to Mobile Sensor Network (MSN) due to sparse network density and sensor node mobility. In this paper, we propose a utility based data gathering protocol (UDG). The distance utility is used to indicate the closeness between sensor nodes and the sink node, and the activity utility is used to evaluate the ability of sensor nodes acting as relays. UDG combines the distance utility with the activity utility to make routing decisions. It also presents a buffer management scheme based on the utility. Experimental results show that UDG achieves desirable performance with low delivery overhead.

Index terms*:* Mobile sensor network, location prediction, order-*k* Markov chain, distance, activity, utility, data gathering.

I. INTRODUCTION

Mobile sensor network (MSN) is the network for pervasive information gathering. It consists of two types of nodes, i.e., the wearable sensor nodes and the high-end sink nodes [1]. The sensor nodes are attached to people (or other mobile objects), gathering target information and forming a loosely connected mobile sensor network for information delivery. The sink nodes (e.g., mobile phones or personal digital assistants with sensor interfaces) are either deployed at strategic locations with high visiting probability or carried by a subset of mobile entities, serving as sinks to receive data from wearable sensors and forward them to access points of the backbone network.

There are many typical applications of MSN, such as flu virus tracking, air quality monitoring [2], traffic condition monitoring, and wildlife tracking. Due to the sparse network density and the nodal mobility, these applications share a unique characteristic that the network connectivity is extremely low and intermittent. In such environments, message replications are necessary for data delivery in order to achieve high delivery ratio, but replications also increase transmission overheads. Therefore, how to reach a well tradeoff between the data delivery ratio/delay and the delivery overhead is the main problem that should be solved effectively in data gathering. An efficient data gathering scheme tailored for MSN needs to maximize the data delivery ratio with the minimum communication overhead.

To address this tradeoff, quota-based protocols [3, 4, 5, 6] were presented in recent years. These protocols decrease the transmission overhead by restricting the number of replicas, that is, an upper bound on the number of replicas allowed in the network is fixed during message creation. However, quota-based protocols assume that all sensor nodes have the same capability of delivering data, ignoring the difference among their delivery abilities. Compared to the quota-based replication, the selectively replication [7, 8, 9] is a better approach to achieve a well tradeoff between the delivery ratio/delay and the overhead. Because this approach maximizes the data delivery ratio with low delivery overhead by only replicating messages to nodes with high possibilities of delivering data to sink nodes successfully.

This paper proposes a utility based data gathering protocol (UDG), which uses the utility of nodes to evaluate their abilities of forwarding messages to the sink node successfully. On one hand, the distance utility, which can be used to indicate the closeness between the sensor node and the sink node, is acquired through making location predictions on the basis of order-*k* Markov model. On the other hand, the activity utility, which can be used to evaluate nodal abilities of acting as relays, is acquired on the basis of historical statistics. The final utility of the sensor node is the combination of its distance utility and activity utility. Through only replicating messages to nodes with higher utilities and managing the limited buffer of nodes efficiently, UDG achieves desirable network performance with low transmission costs.

The rest of this paper is organized as follows: we review the related work and identify the problems in the existing works in Section II. In Section III, we describe the data gathering scheme of UDG in detail. The buffer management scheme is proposed in Section IV. Section V introduces the experimental environment and analyzes the experimental results. Finally, we conclude this work in Section VI.

II. RELATED WORK

Researches on MSN are motivated mainly by the delay-tolerant network (DTN) [10] and its applications in sensor networks and mobile ad hoc networks. Various approaches have been proposed to address the data gathering problem in DTN, most of them are replication-based. Replication-based protocols insert multiple copies of a message into the network to increase the probability of delivering message successfully. For instance, Epidemic [11] replicates messages at all transfer opportunities, hoping to find a path to the destination. However, as flooding messages at all transfer opportunities, this approach results in the waste of resources and the increase of transmission overheads. This leads to work in developing protocols which can achieve better tradeoff between the delivery ratio and the overhead.

SWIM [3], Spray and Wait/Focus [4, 5] are protocols which restrict the number of replicas in the network. However, these protocols have not taken nodes delivery capabilities into consideration, that is, the replication task assigned to a node does not correspond with its delivery ability in most cases, i.e., the task may be too heavy or too easy for the node. In other words, the difference of delivery abilities of different nodes is neglected by these protocols. This results in poor delivery performances and causes a long delay. In fact, delivery abilities of different nodes are obviously not same since they are attached to different objects.

ZebraNet (ZBR) [12] employs mobile sensor nodes to support wildlife tracking. Considering the characteristic of MSN, ZBR proposes a history-based approach for data gathering. It assesses the delivery ability of a sensor node according to the historical records. More specifically, the records that nodes delivery messages to the sink node directly are used to indicate their history level. However, it is not accurate to evaluate ones delivery ability by only using the simple historical level.

On the basis of that, Wang et.al presents RED and FAD [7], and they use the delivery probability to indicate the likelihood that a sensor node can deliver data messages to the sink successfully. Based on RED, FAD further employs a fault tolerance value to constrain message replications. However, nodes delivery probabilities may become inaccurate after multiple recalculations and multiple transmissions, and the buffer management, which is based on the delivery probability, will become inefficient in this situation.

Although the selectively replication [7, 8, 9, 12, 13, 14, 15] is a good approach to achieve a well tradeoff between the delivery performance and the overhead, most solutions lack comprehensive understanding on the delivery ability of sensor nodes in a realistic environment. For instances, MPAD [13], MTAD [14] and MEDR [15] can only work under the assumption that the future destination or the movement direction is known by sensor nodes before they begin to move. In fact, this assumption is not realistic in many scenarios.

III. DATA GATHERING

In this section, we propose a utility based data gathering protocol (UDG) for MSN. We present our protocol based on the following two aspects.

For one thing, since data need to be delivered to the sink node at last, the closer a sensor node is to the sink node, the more likely it is to deliver messages to the sink node successfully. Especially, when the sink node is within the transmission range of a sensor node, messages can be transmitted from the sensor node to the sink node directly. Therefore, it is reasonable to use the distance between the sensor node and the sink node to indicate their closeness. In other words, the distance can be used to evaluate the possibility that a sensor node can deliver messages to the sink node successfully.

For another, the more actively a sensor node is, the more likely it is to contact many other nodes in the given period. Thus an active node can act as a good relay in data transmission.

According to the above two considerations, UDG combines the distance utility with activity utility to make routing decisions. We will describe UDG in detail in this section.

a. Assumptions

As shown in Figure 1, the whole network area with a radius of *R* is partitioned into *m* adjacent concentric parts termed coronas. All coronas have the same width denoted by ω.The sink node is denoted by C_o, and it is fixed (non -mobile) in the center of all the concentric coronas. The i_{th} corona is denoted by C_i, and C_{i-1} is closer to the sink node than C_i. Let Z denote the set of all coronas, i.e. $Z=\{C_1,C_2,\ldots,C_m\}$, and let S denote the set of serial numbers of all coronas, i.e. $S=\{1,2,..,m\}$.

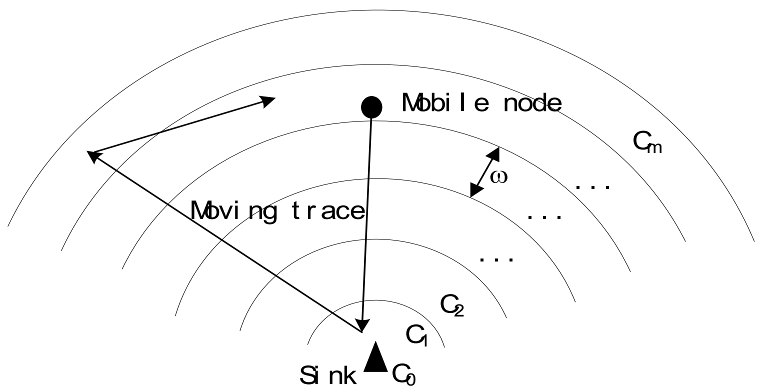

Figure 1. The concentric coronas

Through either GPS (Global Positioning System) or other GPS-less techniques such as those described in [16, 17], each sensor node can compute its physical position.

We assume that the location of the sink node, denoted by (x_s,y_s), is known to all sensor nodes. The current position of sensor node *i* is denoted by (x_i,y_i), and the distance between sensor node *i* and the sink node can be calculated as

$$d_i = \sqrt{(x_i - x_s)^2 + (y_i - y_s)^2} \quad (1)$$

Furthermore, we can get the identifier of the corona which node *i* is currently in,

$$S_{cur} = \left\lceil \frac{d_i}{\omega} \right\rceil \quad (2)$$

where $S_{cur} \in S$. Let $\Delta\tau$ denote the sampling period. Each sensor node obtains its location information at interval of $\Delta\tau$. $\Delta\tau$ can be calculated as

$$\Delta\tau = \frac{\omega}{\bar{v}} \times \lambda \qquad (3)$$

where ω is the width of a corona, and $\bar{v}$ is the average speed of sensor nodes, and λ is a constant ($0<\lambda\leq 1$) , which is used to control the sampling frequency.

b. Location prediction

In many applications, the location information of nodes is useful to improve the network performance. Different approaches of location prediction [18, 19, 20] have been proposed in recent years. In MSN, locations of sensor nodes in the near future are often relevant to their historical locations, therefore, we can use order-k ($O(k)$) Markov chain model to predict nodes locations in the near future. Based on the order-k Markov chain model, the location of a sensor node in the $k+1_{th}$ period can be predicted from the current context, that is, the sequence of the k most recent symbols in the history.

We think of the node location as a random variable X. Let $X(u, v)$ be a string $X_u X_{u+1} \ldots X_v$ representing the sequence of random variables $X_u, X_{u+1}, \ldots X_v$ for any $1 \leq u \leq v \leq m$.

Consider sensor node i, whose location history is $\Omega = l_1, l_2, \ldots, l_j$, where l_j denotes the corona in which node i is lying in the j_{th} sample period, $l_j \in Z$. Let R_k denote the sequence of the latest k location symbols, i.e., $R_k = l_{j-k+1}, l_{j-k+2}, \ldots, l_j$.

Let l denote any a corona, and $l \in Z$. The conditional probability that node i will appear in l in the next period is denoted by $P(X_{j+1} = l \mid X(j-k+1, j) = R_k)$ .The Markov assumption is that X behaves as follows.

$$P(X_{j+1} = l \mid X(1, j) = \Omega) = P(X_{j+1} = l \mid X(j-k+1, j) = R_k) \qquad (4)$$

$$P(X_{j+k+1} = l \mid X(j+1, j+k) = R_k) = P(X_{i+k+1} = l \mid X(i+1, i+k) = R_k) \qquad (5)$$

Equation (4) indicates the assumption that the probability depends only on the context of the k most recent locations. Equation (5) indicates the assumption of a stationary distribution, that is, that the probability is the same anywhere the context is the same.

We can generate an estimate $P(X_{j+1}=l\,|\,X(j-k+1,j)=R_k)$ from the historical location sequences Ω and the current context R_k,

$$P(X_{j+1}=l\,|\,X(j-k+1,j)=R_k)=\frac{Num(R_k l,\Omega)}{Num(R_k,\Omega)} \tag{6}$$

where $Num(R_k l, \Omega)$ is the number of times that the substring $R_k l$ occurs in the string Ω. Notice that if R_k occurs for the first time in the history, then $Num(R_k, \Omega)=1$ and $Num(R_k l, \Omega)=0$. In this situation, the above equation is equal to 0, and the $O(k)$ Markov predictor makes no prediction.

We can solve this problem by gradually reducing the value of k, i.e., $k=k$-1, repeating this process until $O(k)$ Markov predictor can make prediction. In general, whenever an unknown context occurs, the $O(k)$ predictor recursively used the result of the $O(k-1)$ predictor. If $O(1)$ Markov predictor still can't make prediction at last, that means R_1 occurs for the first time in the history, then we can make predictions as

$$P(X_{j+1}=l\,|\,X_j=R_1)=\frac{Num(l,\Omega)}{L_\Omega} \tag{7}$$

where $Num(l, \Omega)$ is the number of times that l occurs in the string Ω, and L_Ω is the length of string Ω.

Figure 2 illustrates the location transitions of a sensor node. Let C_{cur} denote the corona which the sensor node is currently in, and $C_{cur}\in Z$,. C_f is the f_{th} corona which the sensor node may move into in the next period, and $C_f\in Z$. P_f is the possibility that the node may move from C_{cur} to C_f in the next period, and $P_f=P(X_{j+1}=C_f|X(j-k+1,j)=R_k)$, $P_f\geq 0$ and $\sum_{f\in S} P_f = 1$ 。

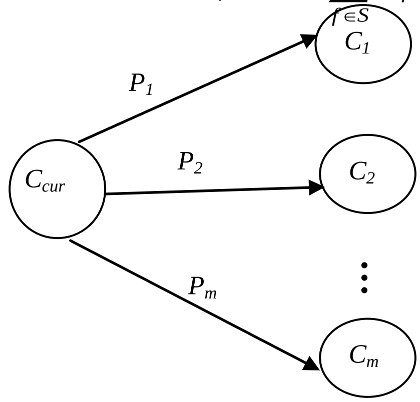

Figure 2. Location transitions

Let S^i_{next} denote the identifier of the predicted corona which node i will move into in the next period. S^i_{next} can be calculated as

$$S^i_{next} = \sum_{f \in S} f \times P_f \tag{8}$$

where f ($f \in S$) is the identifier of corona C_f. The smaller S^i_{next} is, the closer sensor node i will be to the sink node in the next period.

c. Active degree

We use active degree (AD) to indicate the activity of sensor nodes. The active degree of a sensor node is valued by the number of sensor nodes it meets in the period of ΔT. Generally, an active node may meet more other nodes than an inactive one in the given period. In order to evaluate the activity of a node accurately, its historical and current behaviors should be taken into consideration. Therefore we use a time-weighted average of active degrees to indicate the activity of sensor nodes.

Given sensor node i, it obtains its active degree every period of ΔT. Let A_{ij} denote its active degree in the j_{th} period and let t_{ij} be the timestamp of the j_{th} period. Then its average active degree, which is denoted by $\overline{A_i}$, is calculated based on the n active degrees acquired in the most recent n periods,

$$\overline{A_i} = \sum_{j=1}^{n} A_{ij} \times t_{ij} \Big/ \sum_{j=1}^{n} t_{ij} \tag{9}$$

where A_{in} is the active degree obtained in the n_{th} period (the latest period). We can see from Equation (9) that the newly obtained active degree accounts for higher proportions in $\overline{A_i}$, as the recent state of nodes may be a good prediction for the future.

Equation (9) can be further optimized as

$$\overline{A_i} = \frac{\overline{A_i}' \times st_{in-1} + A_{in} \times t_{in}}{st_{in-1} + t_{in}} \tag{10}$$

where $\overline{A_i}'$ is the time-weighted average active degree before it is updated, and st_{in} is the sum of timestamps of the latest n periods.

$$st_{in} = \sum_{j=1}^{n} t_{ij} \tag{11}$$

After optimization, sensor node i only needs to store $\overline{A_i}$ and st_{in} instead of storing n active degrees and n timestamps of the latest n periods.

d. Utility based routing

We use utility to evaluate the ability of delivering data to the sink node successfully for sensor nodes. Let DU_i denote the distance utility of sensor node i compared to sensor node j, and it is given by

$$DU_i = \frac{1 - \frac{S_{next}^i}{m}}{(1 - \frac{S_{next}^i}{m}) + (1 - \frac{S_{next}^j}{m})} \qquad (12)$$

where m is the total number of the concentric coronas (as described in Figure 1). Let AU_i denote the activity utility of sensor node i compared to sensor node j, and it is given by

$$AU_i = \frac{\overline{A_i}}{\overline{A_i} + \overline{A_j}} \qquad (13)$$

Selecting which sensor node as the best carrier for the message becomes a multiple attributes decision problem, where we wish to select the sensor node which provides the maximum utility for carrying the message. The final utility is a value between 0 and 1, and it is the combination of two components: the distance utility and the activity utility. Let U_i denote the final utility of sensor node i, and it is calculated as

$$U_i = DU_i \times \alpha + AU_i \times \beta \qquad (14)$$

where α and β are tunable parameters and $\alpha + \beta = 1$. Consequently these parameters allow for the adjustment of the relative importance of the two utility values.

The data delivery scheme is based on the final utility of nodes. When sensor node i moves into the communication range of a set of L sensor nodes, Let $Y=\{\Psi y \mid 1 \le y \le L\}$ represent the set of L nodes. Sensor node i first compares its utility with those nodes through simple handshaking messages and then replicates messages from itself to a subset of the L sensor nodes, and the members of this subset are those whose final utilities is larger than U_i.

Once a sensor node delivers a message to the sink successfully, this message will not be kept by the sensor node, but its ID will be kept by this node for a short term. Meanwhile, if this node receives this message again, it will reject it and inform the sender to delete it.

IV. QUEUE MANAGEMENT

The replication strategy can be employed to increase the delivery chance. However, the transmission overhead also increases. In this section, we propose a priority value (PV) based queue management approach to address the tradeoff between the delivery performance and the overhead. The queue management is to appropriately sort the data messages in the queue, to determine which data message is to be sent when the sensor node meets another sensor node, and to determine which data message is to be dropped when the queue is full.

Let PV_M^i denote the priority value of message M in the queue of sensor node i. PV_M^i is determined when M is inserted into the queue. For a new message generated from the sensing unit, it is endowed with the highest priority value, i.e., $PV_M^i = H$, and H>0, where H is a constant, indicating the highest priority value. When the message is forwarded, its PV will be updated. For instance, if M is transmitted from node i to node j, PV_M^i will be updated as

$$PV_M^i = PV_M^{i\,'} \times \frac{U_i}{U_j} \tag{15}$$

where $PV_M^{i'}$ is the PV of message M carried by sensor node i before the transmission. From Equation (15) we can see clearly that PV declines in the replication process. The larger U_j is, the more rapidly PV_M^i declines.

The proposed queue management approach is based on the priority value. The delivery order of messages in one's message queue is decided by their PVs, and the possibility of a message being delivered to the sink node can also be reflected by its PV. The message with the highest PV will be send first, while the one with the lowest PV will be dropped when the data queue is full.

More specifically, the message with a larger PV is likely a new message or which has not been delivered to sensor nodes with large utilities, thus the probability that it has been received by the sink node is usually low, and it should be transmitted first when the delivery chance appears. This is done by sorting messages in the queue with a decreasing order of their PVs. The message with the largest PV is always at the top of the queue and should be transmitted first. If the queue is full, the newly arrived message will be inserted into the queue at an appropriate position according to its PV, and the message at the tail of the queue will be dropped. Because this message is the one with the smallest PV, and the possibility that it has been delivered the sink node is higher than that of other messages.

V. EXPERIMENT AND RESULTS

Experiments have been carried out to evaluate the performance of different data gathering approaches. We implement UDG, FAD and ZBR protocols in NS-2.33 and compare performances of the three protocols. We also analyze impacts of different experimental parameters on performances of these protocols.

a. Network model

Since we base our protocol on making location predictions according to the movements of nodes, it is important that the mobility model we use is realistic. Therefore, we use the community model [21], which can reflect the realistic scenario better than some traditional mobility models such as the random waypoint model [22] and the random direction model [23].

In the community model, the network area is divided into several subareas, one of them is called gathering place, and the other areas are called communities. The gathering place may be the center of the town or the feeding ground for animals in realistic scenarios, and the communities may be the village for people or the habitat for animals. Each sensor node belongs to one home community that it is more likely to visit than other places, and for each community there are a number of sensor nodes that have it as the home community.

Each sensor node selects a destination and moves to it at a selected speed. After it arrives at the destination, it pauses there for a while and then repeats the above process. If a sensor node is at home, there is a high possibility for it to select the gathering place as the destination (of course it is also possible for it to go to other places), and if it is away from home community, it is very likely that it will return home.

The mobility of each sensor node follows the process described above and is mutually independent with each other. The sink node is located in the center of the gathering place, and it is fixed (non -mobile).

b. Experimental setup

Without loss of generality, we consider a network that consists of N sensor nodes randomly deployed in a square area A. Each sensor node has the same transmission radius of r. As shown in Figure 3, the network area is divided into 3×3 subareas. Without loss of generality, the bottom right corner area is appointed as the gathering place, and the sink node is located in the center of this area. We assume the data generation of each sensor node follows a Poisson process. The experimental parameters and their default values are summarized in Table 1.

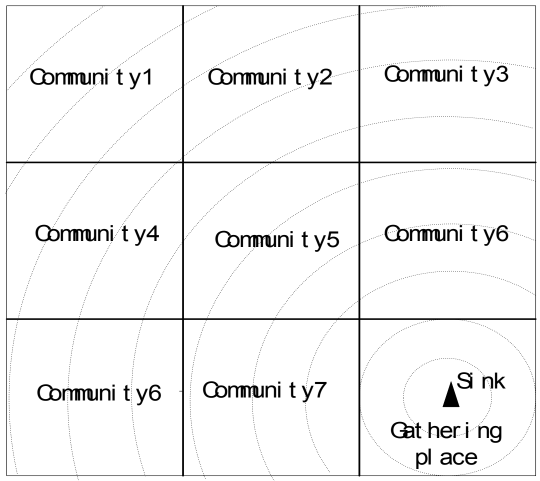

Figure 3. Network model

Table 1. Experimental parameters

Parameter	Default Value
Network size (m^2)	1500×1500
Number of sensor nodes	200
Transmission radii r (m)	50
Queue length of sensor node(message)	200
Speed of sensor node (m)	1-10
Size of data message(bit)	200
Message generation rate (message/s)	0.01
Length of Ω	400
Maximum delay tolerant value t_{max} (s)	1800
k of Markov chain	4
ω (m)	50
α,β of UDG	0.5, 0.5
ΔT of UDG (s)	120
Δ of FAD	10
α of FAD	0.1

c. Experimental results

c.i Impact of queue length

We vary the maximum queue length in our experiments to evaluate the performance of different protocols, with results presented in Figure 4 and Figure 5.

With the increase of queue length, the delivery ratio of all approaches increases, because the number of messages dropped by nodes due to the overflow of the queue declines. As shown in Figure 4, UDG outperforms FAD and ZBR. As expected, ZBR has the lowest delivery ratio, because it only employs history records of direct contacts to decide message transmission. For sensor nodes that never directly meet the sink, data transmissions become random, and thus ZBR is less efficient.

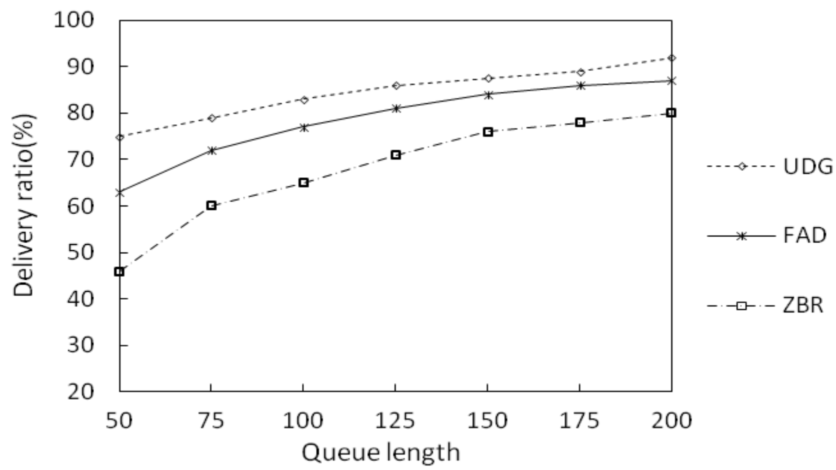

Figure 4. Impact of queue length on delivery ratio

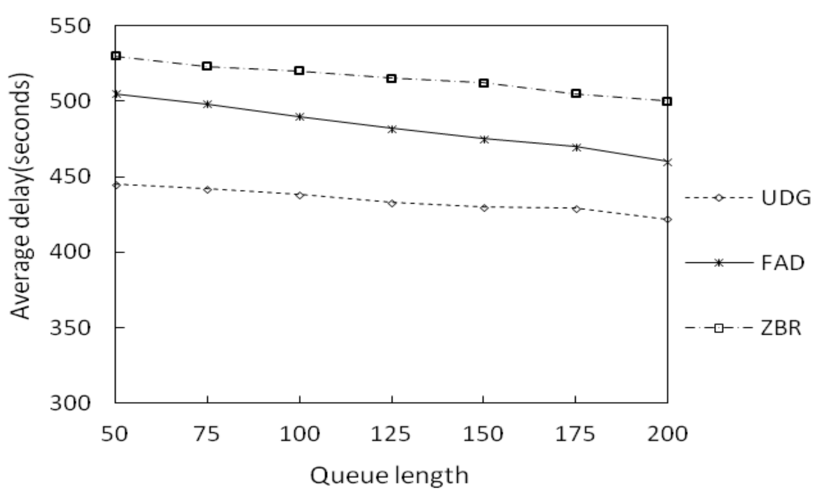

Figure 5. Impact of queue length on average delay

We also vary the queue length to observe its impact on the average delay. As shown in Figure 5, the increase of the queue size has slightly influence on the delay. With the queue length increasing, nodes can buffer more messages, and some messages can be delivered to the sink node more rapidly than before, thus the average delivery delay declines. UDG has a shorter average delay than FAD and ZBR, since it employs the comprehensive utility of nodes to delivery data and to manage the queue efficiently. As expected, ZBR has the longest delivery delay among these three protocols. This is because sensor nodes can raise their historical levels only when they meet the sink node directly. For nodes which never directly meet the sink node, data transmissions become random, thus the delivery delay is long.

c.ii Impact of transmission loads

In this section, we evaluate the performance of the three protocols under different transmission loads by generating messages at different rates. With the rate of message generation changing from 0.01 message/s to 0.1 message/s, the performance of these protocols is shown in Figure 6 and Figure 7.

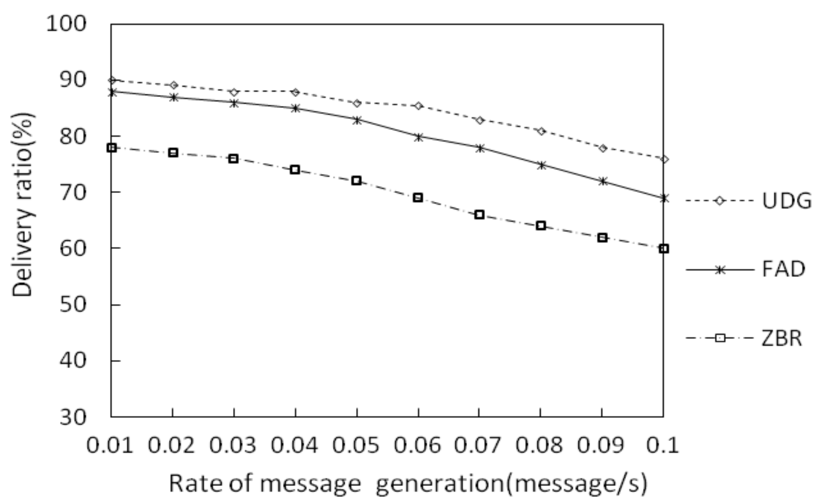

Figure 6. Impact of transmission loads on delivery ratio

We can see from Figure 6 that with the rate of message generation increasing, the delivery ratio declines. The delivery ratio of the three protocols is high when the rate of message generation is low. However, as the rate of message generation increases, network collisions and resource consumptions aggravate, which results in the decline of the delivery ratio. Among the three protocols, UDG achieves the best performance. This means that UDG can deal with heavier transmission loads and has less resource consumption than the other two protocols, since it only replicates messages to nodes with higher utilities, and it manages the queue according to message priorities. For ZBR, as data may be forwarded among nodes randomly, the delivery ratio is lower than the other two protocols.

Figure 7 shows the impact of transmission loads on the average delivery delay. The average delivery delay of all protocols is going up with the rate of message generation increasing. This is because network collisions and resource consumptions are intensified with the increase of transmission loads.

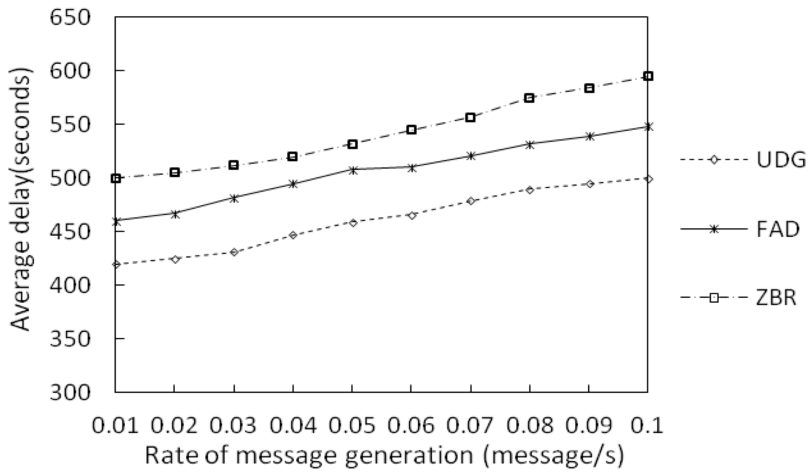

Figure 7. Impact of transmission loads on delay

c.iii Impact of density of nodes

The network connectivity is closely related to the density of sensor nodes. As the network area is changeless, the change of the density of nodes can be reflected by the change of the total number of nodes. Thus, in this section, we will evaluate the impact of the density of nodes on network performances by changing the number of nodes.

We can see from Figure 8 and Figure 9 that UDG performs better in delivery ratio and delivery delay than the other two protocols. As shown in Figure 8, with the number of nodes increasing, the delivery ratio of the three protocols increases at first, however, it decreases when the number of nodes continues to increase. On the contrary, we can see from Figure 9 that with the increase of the number of nodes, the delivery delay of all the protocols decreases at first, but it increases later. The reason of those changes is that the network connectivity is enhanced with the increment of the density of nodes at first, and thus there is a marked improvement in performances of the three protocols. However, with the density of nodes continuing to increase, network collisions and resource consumptions aggravate, which results in the decrease of the delivery ratio and the increase of the delay. Although the influence of the density of nodes on the performance of ZBR is slighter than UDG and FAD, ZBR's performance is inferior to these two protocols. This is because data delivery becomes random when nodes can't meet the sink node directly, which results in the delivery chance in ZBR is less.

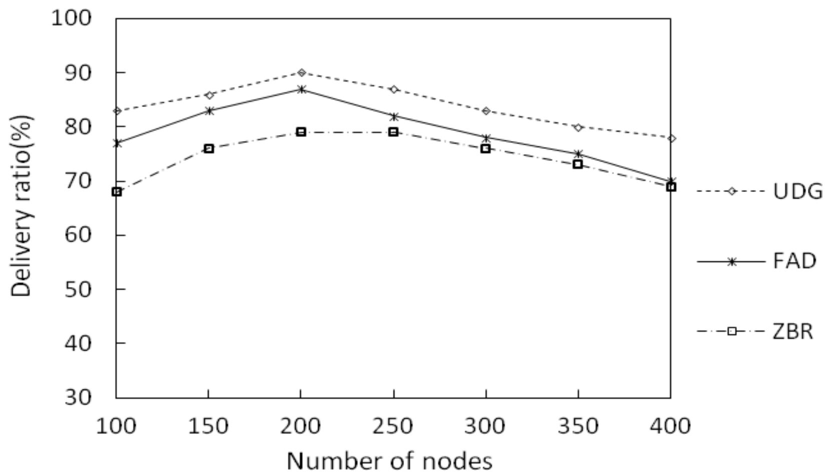

Figure 8. Impact of the density of nodes on delivery ratio

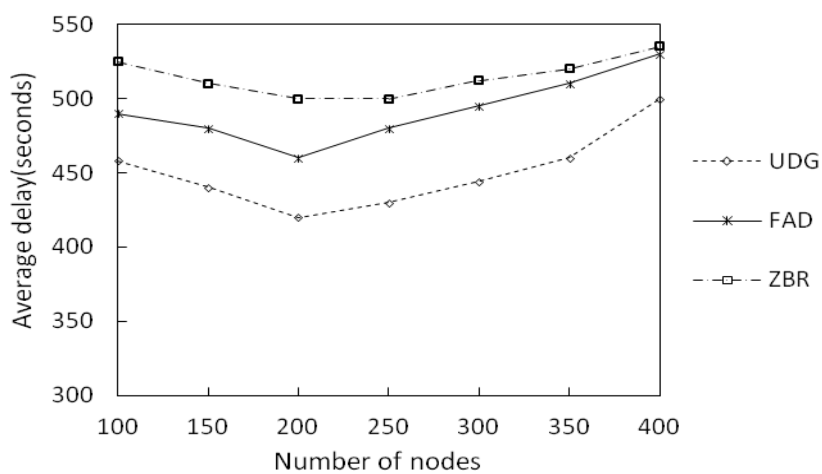

Figure 9. Impact of the density of nodes on delay

c.iv Delivery cost

We compare the total delivery cost of the three different protocols. The total delivery cost is measured by the total delivery number of messages (including duplications) transmitted across the network. We normalize the total delivery cost by dividing it by the total number of unique messages.

As shown in Figure 10, the delivery cost for all protocols increases with the increase of the queue length. This is because nodes can keep more messages when the size of the queue increases, then messages can be transmitted among nodes for more times, and the transmission overhead increases.

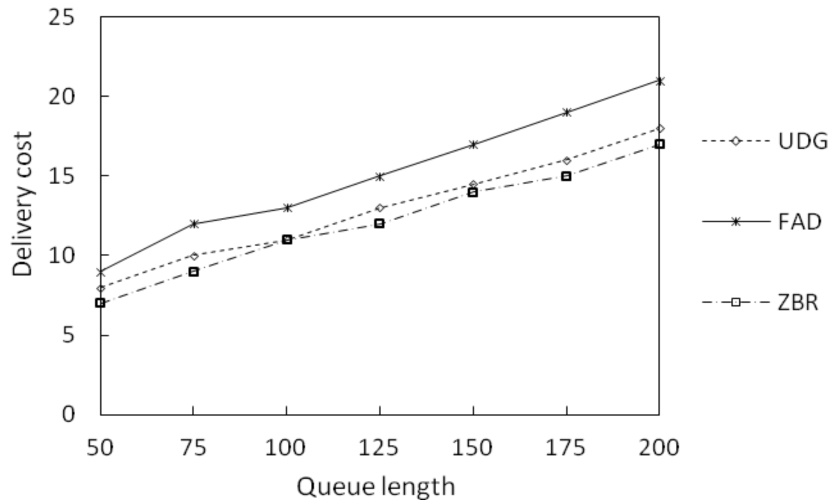

Figure 10. Impact of queue length on delivery cost

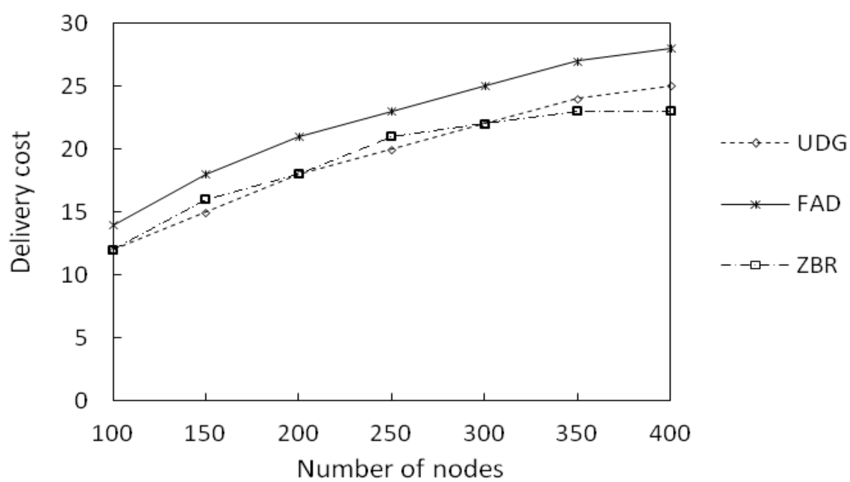

Figure 11. Impact of the density of nodes on delivery cost

We can see from Figure 11 that the delivery cost for all protocols increases with the increase of the density of nodes. The reason is that when the density of nodes increases, the connectivity of the network is improved, thus there are more chances for nodes to transmit messages, and the delivery overhead increases. Among these protocols, UDG's delivery cost is lower than FAD. Although the delivery overhead of ZBR is comparable to that of UDG, its delivery ratio and delay are not desirable. The reason is that transmission chances in ZBR are less than those in other two protocols, and message transmissions become random when nodes can't meet the sink node directly.

VI. CONCLUSIONS

In the intermittent connected environment of MSN, the strategy of selectively replication is necessary for data gathering to achieve a well tradeoff between the network performance and transmission costs. This paper proposes a utility based data gathering protocol (UDG) for MSN. UDG combines the distance utility with the activity utility of sensor nodes to make routing decisions and to manage the limited buffer efficiently. Experimental results show that UDG achieves better delivery performance with lower transmission overhead than other two relevant schemes.

ACKNOWLEDGMENTS

This paper is supported by the National Natural Science Foundation of China under Grants No. 61202444, 61103226 and 60903158, the Fundamental Research Funds for the Central Universities of China under Grant No.ZYGX2011J102 and the National High Technology Research and Development Program of China under Grant No.2012AA041403.

REFERENCES

[1] Y. Wang, F. Lin and H. Wu, "Poster: Efficient Data Transmission in Delay Fault Tolerant Mobile Sensor Networks", IEEE International Conference on Network Protocols, pp.1021-1034, November 2005.

[2] J. H. Liu, Y. F. Chen, T. S. Lin, C. P. Chen, P. T. Chen, T. H. Wen, C. H. Sun, J. Y. Juang and J. A. Jiang, "An Air Quality Monitoring System for Urban Areas Based on the Technology of Wireless Sensor Networks", International Journal on Smart Sensing and Intelligent Systems (S2IS), Vol.5, No.1, pp.191-214, 2012.

[3] T. Small and Z. J. Haas, “The Shared Wireless Infostation Model: A New Ad Hoc Networking Paradigm”, The 4th ACM International Symposium on Mobile Ad Hoc Networking, pp.233−244, June 2003.

[4] T. Spyropoulos, K. Psounis and CS. Raghavendra, “Spray and Wait: An Efficient Routing Scheme for Intermittently Connected Mobile Networks”, ACM SIGCOMM Workshop on Delay Tolerant Networking, pp 252−259, August 2005.

[5] T. Spyropoulos, K. Psounis and CS. Raghavendra, “Spray and Focus: Efficient Mobility-Assisted Routing for Heterogeneous and Correlated Mobility”, Proceeding of IEEE PerCom Workshop on Intermittently Connected Mobile Ad Hoc Networks, pp.902-910, March 2007.

[6] X. J, F. X and W. R. Chuan, “Adaptive Spray Routing for Opportunistic Networks”, International Journal on Smart Sensing and Intelligent Systems (S2IS), Vol.6, No.1, pp.95-119, 2013.

[7] Y. Wang and H. Y. Wu, “Delay/Fault-Tolerant Mobile Sensor Network (Dft-Msn): A New Paradigm for Pervasive Information Gathering”, IEEE Transactions on Mobile Computing, Vol.6, No.9, pp.1021-1034, September 2007.

[8] J. Y. Liu, M. Liu, H. G. Gong and J. Z. Zeng, “Expected Shortest Path Routing for Social-oriented Intermittently Connected Mobile Network”, Journal of Convergence Information Technology, Vol.7, No.1, pp.94-101, January 2012.

[9] J. Wu and Y. S. Wang, “Social Feature-based Multi-path Routing in Delay Tolerant Networks”, Proceeding of IEEE INFOCOM, pp.1368-1376, March 2012.

[10] K. Fall, “A Delay Tolerant Networking Architecture for Challenged Internets”, International Conference on ACM Special Interest Group on Data Communication, pp.34-45, August 2003.

[11] A. Vahdat and D. Becker. “Epidemic Routing for Partially Connected Ad Hoc Networks”, Technical Report, CS-2000-06, Duke University, 2000.

[12] P. Juang, H. Oki and Y. Wang, “Energy-Efficient Computing for Wildlife Tracking: Design Tradeoffs and Early Experiences with Zebranet”, The 10th International Conference on Architectural Support for Programming Languages and Operating Systems, pp.96−107, October 2002.

[13] J. Q. Zhu, J. N. Cao, M. Liu, Y. Zhen, H. G. Gong and G. H. Chen, “A Mobility Prediction-based Adaptive Data Gathering Protocol for Delay Tolerant Mobile Sensor Network”, International Conference on GLOBECOM, pp.1-5, November 2008.

[14] F. L. Xu, M. Liu, J.N. Cao, G. H. Chen, H. G. Gong and J. Q. Zhu, “A Motion Tendency-based Adaptive Data Delivery Scheme for Delay Tolerant Mobile Sensor Networks”, International Conference on GLOBECOM, pp.1-6, November 2009.

[15] Y. Feng, M. Liu, X. M Wang and H. G. Gong, “Minimum Expected Delay-based Routing Protocol (MEDR) for Delay Tolerant Mobile Sensors Networks”, Sensors, Vol.10, No.9, pp.8348-8362, September 2010.

[16] X. Cheng, D. Z. Du and L. Wang, “TPS: A Time-Based Positioning Scheme for Outdoor Sensor Networks”, Proceedings of IEEE INFOCOM, pp. 7–11, March 2004.

[17] A. Thaeler, M Ding and X. Z. Cheng, “ITPS: An Improved Location Discovery Scheme for Sensor Networks with Long Range Beacons”, Journal of Parallel and Distributed Computing, Vol.65, No.2, pp. 8–106, February 2005.

[18] A. Bhattacharya and S. Das, “Lezi-Update: An Information Theoretic Approach to Track Mobile Users in Pcs Networks”, ACM/KLUWER Wireless Networks, Vol.8, No.2, pp.121–135, March 2002.

[19] C. Cheng, R. Jain and E. V. D. Berg, “Location Prediction Algorithms for Mobile Wireless Systems”, Wireless Internet Handbook, pp. 245-263, 2003.

[20] L. B. Song, D. Kotz and R. Jain, “Evaluating Location Predictors with Extensive Wi-Fi Mobility Data”, Proceedings of INFOCOM, Vol. 2, pp.1414-1424, March 2004.

[21] A. Lindgren, A. Doria and O. Schelen, “Probabilistic Routing in Intermittently Connected Networks”, Lecture Notes in Computer Science, Vol. 3126, pp.239–254, August 2004.

[22] B. Christian, H. Hannes and P. C. Xavier, “Stochastic Properties of the Random Waypoint Mobility Model”, Wireless Networks, Vol.10, No.5, pp.555-567, September 2004.

[23] C. Bettstetter, “Mobility Modeling in Wireless Networks: Categorization, Smooth Movement, and Border Effects”, ACM SIGMOBILE Mobile Computing and Communications Review, Vol.5, No.3, pp.55−66, July 2001.