A New Data Gathering Scheme Based on Set Cover Algorithm for Mobile Sinks in WSNs

Yutaro Sasaki†, Hidehisa Nakayama†, Nirwan Ansari‡, Yoshiaki Nemoto†, and Nei Kato†
†Graduate School of Information Sciences, Tohoku University, Japan
‡Advanced Networking Laboratory, ECE Department, NJIT, USA
E-mail: keydaimon@it.ecei.tohoku.ac.jp

Abstract—**Recent Advances in solid state and packaging technologies have enabled production of more efficient and reasonably small devices such as Micro Electro Mechanical Systems (MEMS). Wireless Sensor Networks (WSNs) can gather data from sensor nodes and are now at the practical stage of realizations because of the above advances. Conventional researches have mainly focused on extending the lifetime of WSNs because sensor nodes are only equipped with small-capacity batteries. Mobile Ubiquitous LAN Extension (MULE) is one of the approaches to meet such demand, and it can gather data from isolated nodes. The KAT mobility scheme is one of the mobility schemes on MULE focusing on the efficiency. Therefore, this scheme is expected to prolong the lifetime of the network. However, this scheme cannot ensure that the mobile sinks can gather the data from all of the nodes. In this paper, we focus on the fairness issue of data gathered by the mobile sinks while also considering the efficiency of data gathering. We propose a new mobility scheme based on a new clustering method and the Set Cover Algorithm to ensure that the mobile sinks can gather data from all of the nodes, and simulation results show that fairness of data gathering by the proposed mobility scheme is greatly improved as compared to conventional KAT mobility scheme.**

I. INTRODUCTION

The enabling technology for producing small, reasonable, and high efficiency devices such as Smart Dust [1] has facilitated construction of large networks consisting of many sensor nodes. Data collection from sensor nodes using the standard protocol such as Zigbee, where the transmission range of a sensor node is short, has been performed by multi-hop communications.

Using a multi-hop scheme with fixed sinks as in the conventional approach exhibits some problems. First, the sinks may not be able to gather sensed data from some isolated sensor nodes because they cannot communicate with neighboring nodes. Second, there is an energy problem. In the multi-hop and fixed sink scheme, there is tremendous volume of multi-hop traffic to communicate with long-distant nodes, whereas traffic in the mobile sink scheme is predominantly single-hop traffic. In general, the lifetime of the network in the fixed sink scheme is shorter than that with the mobile sink scheme because the mobile sinks are equipped with rechargeable batteries. Consequently, the mobile sink scheme can reduce the required communication energy. In addition, the mobile sink scheme enables nodes to communicate with all of the nodes because the mobile sink can move anywhere. Therefore, we adopt the MULE [2] approach which uses the mobile sink scheme. This concept is illustrated in Fig. 1 as a three-tier architecture. As mentioned earlier, a node, which can move around the deployed nodes, is referred to as a mobile sink. It moves around the deployed nodes, receives data from them, and sends data to access points. Among various methods which have adopted the mobile sink concept, the KAT mobility scheme [3] has been demonstrated to offer a promising solution for data gathering. It uses at least one mobile sink to gather the sensed data by constructing clusters by using the k-means method [4] and calculating the TSP (Traveling Salesman Problem) path to move around the nodes efficiently.

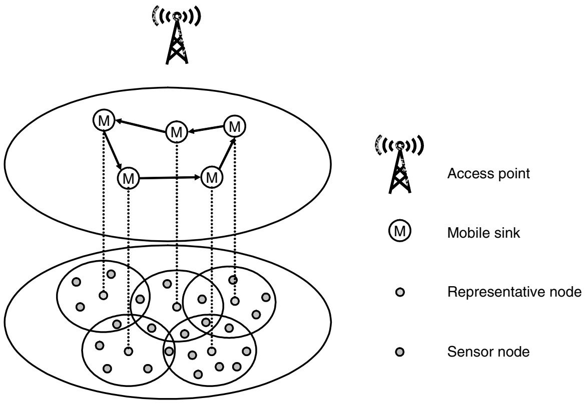


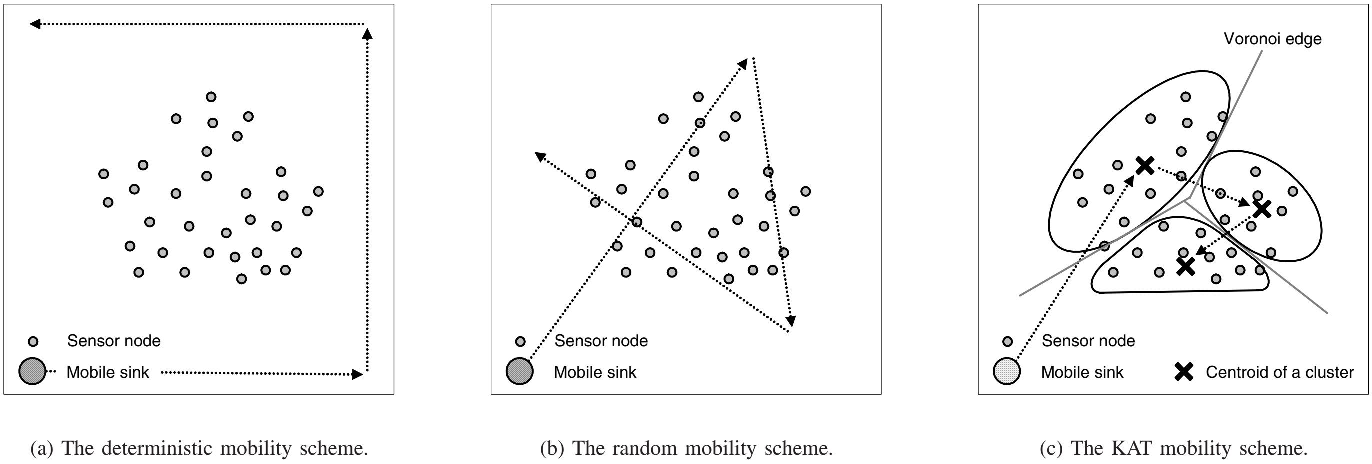
Fig. 1. The concept of the three-tier architecture.

Whereas we can efficiently gather data, it is also desired to obtain data fairly. In this paper, we propose a new mobility scheme by using a new clustering scheme and the Set Cover Algorithm (*SCA*) to mitigate the drawbacks of the conventional data collection scheme. The remainder of this paper is organized as follows. Section II reviews related works including the KAT mobility scheme. Section III presents the proposed algorithm. Section IV presents and discusses the simulation results. Finally, conclusion is described in Section V.

II. RELATED WORKS

Many research works such as LEACH [5] and PEGASIS [6] have adopted the static sink scheme. In this paradigm, there are three mainstreams: cluster-based, tree-based, and multi-path approaches. The representative works of cluster-based, tree-based and multipath approaches are Cougar [7], TAG [8] and

978-1-4244-2324-8/08/$25.00 © 2008 IEEE.



(a) The deterministic mobility scheme.
(b) The random mobility scheme.
(c) The KAT mobility scheme.

Fig. 2. The overview of each mobility scheme.

Synopsis Diffusion [9]. However, many research works have recently adopted the mobile sink scheme because the mobile sink scheme can reduce the energy consumed by the deployed sensor nodes. Thus, we focus only on the mobile sink scheme in this work for the reason justified above. In particular, we focus on how the mobile sinks collect data from the nodes which are randomly distributed.

Many works have been reported on gathering data by using mobile sinks. In evaluating the efficiency of data gathering and collection in WSNs, throughput per unit of power is a more important performance metric than data throughput per unit of time because WSNs are engineered to save their consumed energy to prolong their lifetime. Hereunder, we introduce some conventional path calculation methods. Fig. 2 illustrates each respective mobility scheme. In this figure, small filled circles, big filled circle, closed circular boundaries, and dotted lines denote the sensor nodes, the mobile sink, the clusters, and the trajectory of the mobile sink, respectively.

First of all, the deterministic waypoint mobility scheme is a very simple method in which the mobile sink moves along the arbitrary and unicursal path deterministically. Fig. 2(a) shows the image when a mobile sink moves along the edges of a sensed area deterministically. Secondly, according to the so-called the random waypoint mobility scheme [10], mobile sink decides their next destination randomly and moves toward it. Since the next destination is chosen randomly, it is the safest way against many kinds of attacks such as physical capturing or jamming attack. Fig. 2(b) illustrates the trajectory of a mobile sink according to the random waypoint mobility scheme. Thirdly, the KAT mobility scheme, which was proposed recently, divides all deployed nodes into k clusters by using the k-means clustering algorithm, calculates the TSP path among the centroids of all of the clusters, and navigates the mobile sink along the calculated path. Fig. 2(c) illustrates the trajectory of a mobile sink traversing among the centroids of created clusters.

It was shown that the KAT mobility scheme is more efficient than the other two conventional ones in terms of the throughput per unit of power [3].

Note that these three mobility schemes have a common shortcoming, i.e., there may be some nodes from which the mobile sink cannot gather sensed data for a certain period. In the deterministic mobility scheme, there may be some nodes from which sinks are unable to gather data because the algorithm decides the arbitrary path such as the rectangle path illustrated in Fig. 2. In the random waypoint mobility scheme, the path decided by the algorithm is entirely random, and so it is likely that the mobile sink cannot communicate with all of the nodes. In the KAT mobility schemes, initial deployment of nodes and the given parameter k decide the path deterministically. So we can think of the KAT mobility scheme as one of the deterministic mobility schemes. Therefore, some nodes may not communicate with other nodes in a cluster under certain conditions described below. In this paper, we focus on this aspect and propose a new scheme which aims to ensure that the mobile sink can gather sensed data from all of the sensor nodes.

III. PROPOSED DATA GATHERING SCHEME

First of all, we assume that there are many sensor nodes which can sense environmental data around them in the field. Since these sensor nodes are deployed randomly, some areas are sparse and other areas are dense. Although the KAT mobility scheme aims to gather data efficiently, it exhibits some problems. More concretely, the mobile sink cannot gather data from certain nodes when the number of clusters k or the maximum hop count is small. Under these conditions, the communication range from the centroid point does not correspond to the range of the cluster, and so the mobile sink cannot always communicate with all of the nodes in the cluster when the mobile sink reaches the centroid of the cluster.

In this section, we introduce a new clustering method and SCA (Set Cover Algorithm) to efficiently gather data in WSNs.

A. Clustering Algorithm

In our proposed algorithm, sensor nodes are partitioned into clusters by forming the set of nodes to which the ith node can communicate with as the ith cluster. The ith node is assigned as the ith cluster head since only the ith node is guaranteed to communicate with all of the nodes in the ith cluster. Therefore, the cluster head is usually located near the centroid of its cluster. Actually, we we can form these clusters by using the information of the physical communication range, the location of each node, and the maximum hop counts of communication. However, we formed these clusters by actually observing the nodes to which each node can communicate with.

B. Set Cover Algorithm

The SCA can solve the SCP (Set Cover Problem), which is one of the oldest and most studied NP-hard problems. The SCA produces the set of clusters covering all of the sensor nodes with the least number of clusters. Denote IC as the set of all of the clusters. Therefore, IC is a family of sets. There are many different combinations of sets of clusters that can cover all of the sensor nodes. The exact SCA produces a set of clusters covering all of the sensor nodes with the least number of clusters. However, this optimization problem is NP-Hard [11]. Therefore, we adopt the greedy algorithm which yields an approximation ratio of $\ln |m|$ [12] to reduce the calculation cost, where m is the largest size of the input sets.

Consequently, greedy SCA can be summarized as follows. The input of the algorithm is IC; the algorithm iteratively and heuristically picks the largest cluster, and then the resulting members of the cluster are recognized as covered sensor nodes. Here, the largest cluster is referred to the one which contains the largest number of uncovered sensor nodes. We denote the set of clusters produced by this algorithm as OC, and it is also a family of sets like IC.

C. Proposed Procedure

The overview of the proposed mobility scheme is illustrated in Fig. 3. In this scheme, the algorithm first forms the clusters, and then refines the clusters from all of the clusters by using SCA. Finally, the algorithm calculates the TSP path among the cluster heads, and the mobile sink moves along the path. The procedure can be visualized in Fig. 3 (a)–(c). The details are described below. Initially, the number of clusters equals to the number of nodes n as visualized in Fig. 3 (a).

Step 1. The SCA computes OC. Hereafter, the algorithm only considers all of the clusters in OC, which is a subset of IC.

Step 2. The TSP path among all of the cluster heads in OC is calculated by using the TSP algorithm.

Here, we adopt the approximation algorithm 2-Opt and Or-opt instead of the exact TSP algorithm to reduce the computational costs since TSP is NP-hard [15]. Thus, readers are referred to References [13], [14] for details.

Consider the extreme situation in which each node cannot communicate with any other nodes because the distribution of nodes is too sparse. Even in this case, the proposed algorithm enables the mobile sink to gather all of the sensed data without special setting because we can regard each node as a cluster in this case. To show the superiority of this scheme, we have conducted the simulation as described in the next section.

IV. PEFORMANCE EVALUATION

Qualnet 3.9.5 is adopted as our simulation platform. In this simulation environment, the mobile sink moves to the destination and gathers data for a given time interval via multi-hop communication. Note that the mobile sink gathers data not only when they pause but also when they are in transit to the next destination. The destinations of the proposed mobility scheme are decided by the clustering algorithm and SCA. As stated in Section III-A, this clustering algorithm is based on the communication range.

A. Simulation Setup

1) Network Model: Directed Diffusion [16], which is the most popular protocol stack, is adopted in our simulations. Especially, we adopt the One Phase Pull model [17] in this protocol stack because it is the simplest version of Directed Diffusion. The link layer and the physical layer correspond to those of IEEE 802.11b, which are the global standard used in wireless sensor networks.

2) Battery Model: We adopt LCP (Load current profile) based battery model which is widely used in sensor network simulations. It is assumed that this model is approximated by an N-step staircase function.

B. Performance Metrics

Firstly, we evaluate the fairness by using the following Fairness index F [18]:

$$F = \frac{\left(\sum_{i=1}^{n} d_i\right)^2}{n \sum_{i=1}^{n} {d_i}^2} \tag{1}$$

where d_i denotes the data received by the ith sensor node. Note that this index is equal to 1 when the data from all of the nodes are equivalent, and is close to 0 when the variance of all data is large.

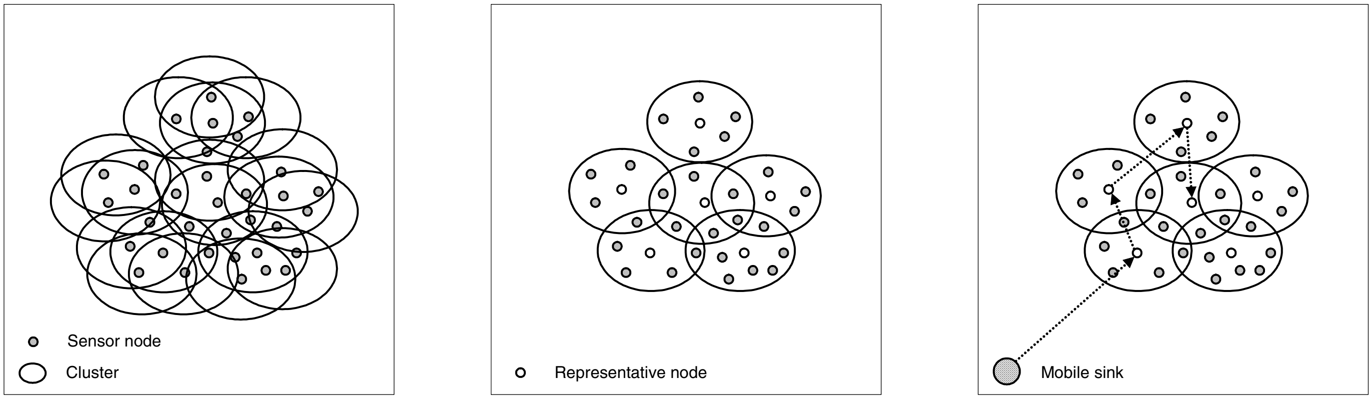
Secondly, we adopt the efficiency index defined below because this metric considers the tradeoff between the throughput and energy consumption [3].

$$R \quad [\text{KB}] = \frac{\text{Received Bytes by all mobile sinks}}{N \times M} \tag{2}$$

$$C \quad [\text{mWhr}] = \frac{\text{Consumed Energy by all sensors}}{M} \tag{3}$$

$$E \quad [\text{KB/mWhr}] = \frac{R}{C} \tag{4}$$

The parameters N and M denote the number of nodes and the number of mobile sinks, respectively. We adopt the above efficiency to validate the performance of our proposed scheme by using the Qualnet simulator.



(a) There are n clusters.

(b) Chosen clusters by the algorithm.

(c) The trajectory along the TSP path.

Fig. 3. The overview of the proposed mobility scheme.

C. Parameter Settings

The simulation time is fixed to 80 minutes, and the area is 5000×5000. The number of mobile sinks is set to 1, its pause time after having arrived at the destination is set to 20 seconds, and its velocity varies from 20 m/s to 30 m/s randomly. The number of sensor nodes varies from 20 to 200. The number of clusters in the KAT mobility scheme is set to 10. In addition to the parameters described above, we adopt the same simulation parameters settings in [3]. We conducted the simulations with ten different seeds for each set of nodes, and the results were averaged.

D. Experimental Results

Fig. 4 shows that the proposed mobility scheme outperforms the KAT mobility scheme in terms of the fairness of the gathered data. Besides, Fig. 5 shows that the proposed mobility scheme exceeds the KAT mobility scheme in terms of the efficiency of data gathering when the number of nodes is small, whereas the proposed mobility scheme achieves at least the same efficiency as that of the KAT mobility scheme when the number of nodes is large.

In the proposed mobility scheme, the algorithm decides the shortest path among the paths where the mobile sink traverses all cluster heads chosen by SCA. In other words, the mobile sink can quickly acquire the data from all of the nodes.

Figs. 6 and 7 show the results of the total number of bytes received from all nodes. Each point of the curve represents one of the simulation results where the number of nodes is 200. In these graphs, we can see that the proposed mobility scheme can gather data from all scattered nodes, whereas data from many nodes cannot be acquired by the mobile sink in KAT.

V. CONCLUSION

In this paper, we have proposed a new mobility scheme by using a new clustering method, which is based on the communication range of each node, to gather data from all of the sensor nodes. Then, the clusters, which cover all of the sensor nodes with the least number of clusters, are further selected by the set cover algorithm. Finally, the algorithm solves the traveling salesman problem to determine the path of the mobile sink, and it navigates the mobile sink to the destination along the TSP-path among the cluster heads. The objective of this mobility scheme is to achieve both higher fairness of gathered data and higher efficiency of data gathering, thus ensuring that the mobile sink can gather data from all of the nodes. Therefore, this scheme can even gather data from the nodes, which are isolated from the neighboring nodes.

Simulation results show that the proposed mobility scheme acquires higher fairness than that of the KAT mobility scheme, and exceeds or achieves the same efficiency as that of the KAT mobility scheme, which has achieved higher efficiency than conventional ones. Therefore, we can conclude that the proposed algorithm can gather data from all of the nodes fairly without compromising efficiency.

ACKNOWLEDGEMENTS

This work was supported through the Strategic International Cooperative Program, JST and NSF (under grant no. 0726549).

REFERENCES

[1] J. Kahn, R. Katz, and K. Pister, "Next century challenges: mobile networking for smart dust," in *Proc. of ACM/IEEE Int. Conf. on Mobile Comput. and Netw.*, Seattle, USA., Aug. 1999, pp. 271–278.

[2] R. Shah, S. Roy, S. Jain, and W. Brunette, "Data mules: modeling and analysis of a three-tier architecture for sparse sensor networks," *Ad Hoc Networks*, vol. 1, no. 2–3, pp. 215–233, Sep. 2003.

[3] H. Nakayama, N. Ansari, A. Jamalipour, and N. Kato, "Fault-resilient sensing in wireless sensor networks," *Comput. Commun. archive*, vol. 30, no. 11–12, pp. 2375–2384, Sep. 2007.

[4] K. Fukunaga, *Statistical Pattern Recognition*, 2nd Ed., Academic Press, 1990.

[5] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 660–670, Oct. 2002.

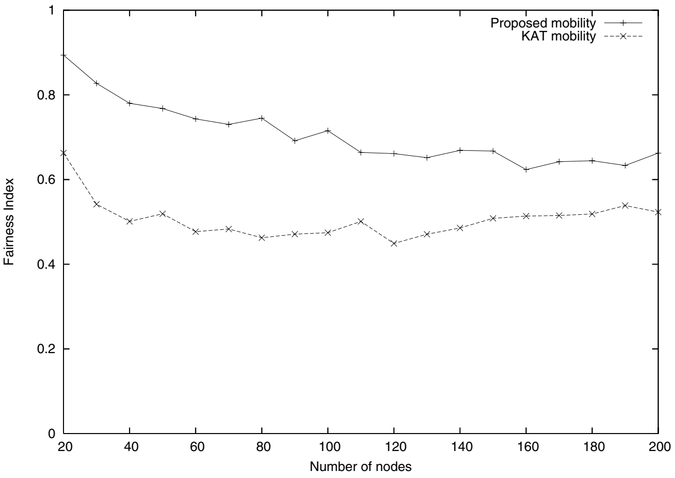


Fig. 4. Fairness of the two respective mobility schemes.

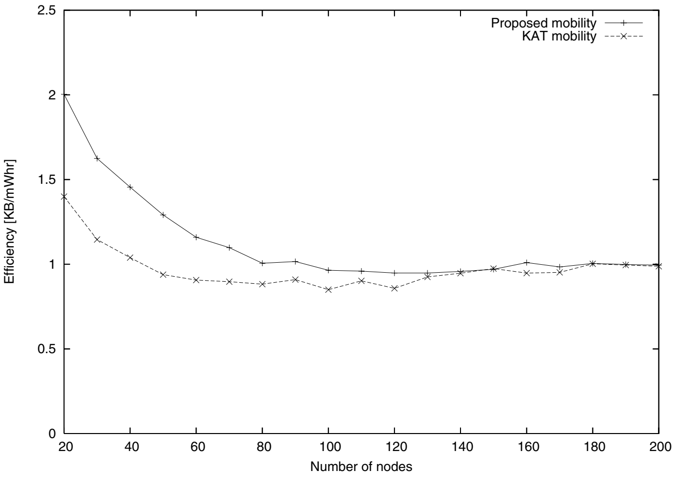


Fig. 5. Efficiency of the two respective mobility schemes.

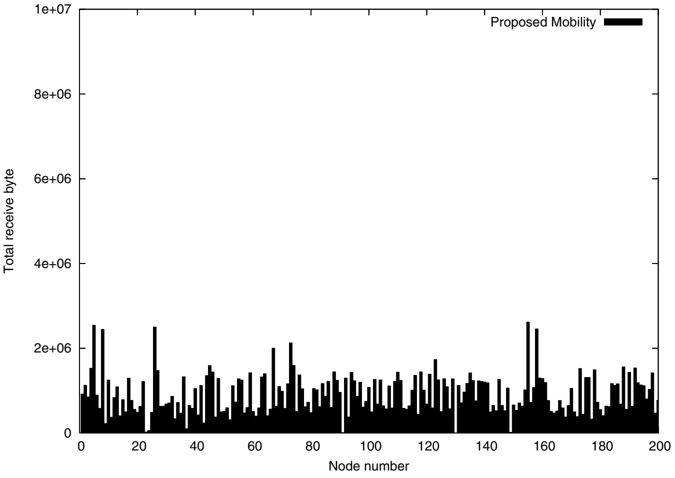


Fig. 6. Total received bytes by the proposed mobility scheme.

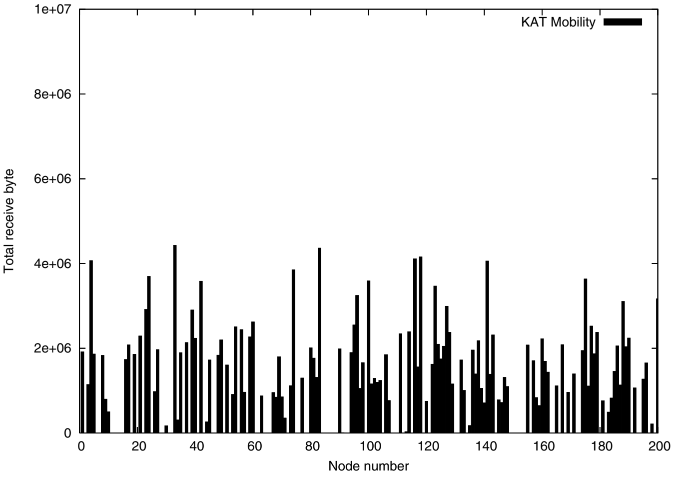


Fig. 7. Total received bytes by the KAT mobility scheme.

[6] S. Lindsey, C. Raghavendra, and K. Sivalingam, "Data gathering algorithms in sensor networks using energy metrics," *IEEE Trans. Parallel Distrib. Syst.*, vol. 13, no. 9, pp. 924–935, Sep. 2002.

[7] Y. Yao and J. Gehrke, "The Cougar Approach to In-Network Query Processing in Sensor Networks," *ACM SIGMOD*, vol. 31, no. 3, pp. 9–18, Sep. 2002.

[8] S. Madden, *et al.*, "TAG: a Tiny AGgregation Service for Ad Hoc Sensor Networks," in *Proc. of OSDI*, Boston, USA., Dec. 2002, pp. 131–146.

[9] S. Nath, P. Gibbons, S. Seshany, and Z. Andersony, "Synopsis Diffusion for Robust Aggregation in Sensor Networks," in *Proc. of ACM SenSys*, Baltimore, USA., Nov. 2004, pp. 250–262.

[10] D. Johnson and D. Malt, "Dynamic source routing in ad hoc wireless networks," *Mobile Computing*, 1996, pp. 153–181.

[11] R. Karp, "Reducibility Among Combinatorial Problems," in *Complexity of Computer Computations*, New York, USA., 1972, pp. 85–103.

[12] D. Johnson, "Approximation algorithms for combinatorial problems," in *Proc. of ACM symp. Theory of comput.*, Austin, USA., 1973, pp. 38–49.

[13] D. Johnson and L. McGeoch, *The Traveling Salesman Problem: A Case Study in Local Optimization*, John Wiley & Sons, 1997.

[14] N. Ansari and E. Hou, *Computational Intelligence for Optimization,* Springer, 1997.

[15] M. Garey and D. Johnson, *Computers and Intractability: A Guide to the theory of NP-Completeness*, *W. Freeman and Company*, 1979.

[16] C. Intanagonwiwat, *et al.*, "Directed diffusion for wireless sensor networking," *IEEE/ACM Trans. Netw.*, vol. 11, no. 1, pp. 2–16, Feb. 2003.

[17] B. Krishnamachari and J. Heidemann, "Application-specific modelling of information routing in wireless sensor networks," in *Proc. of IEEE Int. Conf. on Performance, Comput., and Commun.*, Phoenix, USA., Apr. 2004, pp. 717–722.

[18] R. Jain, D. Chiu, and W. Hawe, "A Quantitative Measure Of Fairness And Discrimination For Resource Allocation In Shared Computer Systems," *DEC Technical Report, DEC-TR-301*, Sep. 1984.