

A Codebook-based Redundancy Suppression Mechanism with Lifetime Prediction in Cluster-based WSN

Huan Chen, Bo-Chao Cheng, Chih-Chuan Cheng, Yi-Geng Chen, and Yu Ling Chou

Abstract—Wireless Sensor Network (WSN) comprises of sensor nodes which are designed to sense the environment, transmit sensed data back to the base station via multi-hop routing to reconstruct physical phenomena. Since physical phenomena exists significant overlaps between temporal redundancy and spatial redundancy, it is necessary to use Redundancy Suppression Algorithms (RSA) for sensor node to lower energy consumption by reducing the transmission of redundancy. A conventional algorithm of RSAs is threshold-based RSA, which sets threshold to suppress redundant data. Although many temporal and spatial RSAs are proposed, temporal-spatial RSA are seldom to be proposed because it is difficult to determine when to utilize temporal or spatial RSAs. In this paper, we proposed a novel temporal-spatial redundancy suppression algorithm, Codebook-based Redundancy Suppression Mechanism (CRSM). CRSM adopts vector quantization to generate a codebook, which is easily used to implement temporal-spatial RSA. CRSM not only achieves power saving and reliability for WSN, but also provides the predictability of network lifetime. Simulation result shows that the network lifetime of CRSM outperforms at least 23% of that of other RSAs.

Keywords—Redundancy Suppression Algorithm (RSA), Threshold-based RSA, Temporal RSA, Spatial RSA and Codebook-based Redundancy Suppression Mechanism (CRSM)

I. INTRODUCTION

In recent years, Wireless Sensor Network (WSN) is employed in many applications, e.g., wild animal tracks [1] and habitat surveillance [2]. Wireless Sensor Network consists of sensor nodes which are distributed in sensor field to measure physical phenomena, such as temperature, humidity, seismic, acoustic etc., and transmits the readings to a fusion center, named as base station (BS), to recover the original data. Due to the limited resource of sensor nodes, the energy efficiency becomes a major challenge of WSN design.

In a sensor field that requires continuous surveillance, grouping sensor nodes into clusters can save energy consumption of sensor nodes because it can reduce the number of transmission and transmission distance of sensor nodes [3]. In a cluster, a cluster head is responsible for relay data to a fusion center for its cluster member. Intuitively, cluster heads represent the best location to perform the data suppression.

Since physical phenomena exists significant overlaps between temporal redundancy and spatial redundancy, there is a need for an intelligent redundancy suppression algorithm

that enables WSN to suppress redundant transmissions while meeting acceptable distortion for reconstruction of original data. Temporal redundancy is due to continuous observation of physical phenomena, while spatial redundancy is due to dense deployment of sensor nodes. The conventional algorithm of RSAs is threshold-based RSA, which can be categorized into three types: (1) Temporal RSA, (2) Spatial RSA, and (3) Temporal-spatial RSA.

The design discipline of temporal RSA is to reduce transmission of temporal redundant data observed by sensor nodes by setting thresholds. In literature, TEEN [4] uses two thresholds T_H and T_S to define different data suppression behaviors. The former is used to check whether a critical event occurred, in which scenario data shall be transmitted without any suppression. The later threshold is a general threshold to guarantee the distortion requirement. It is expected that if a larger threshold is used (which allows more distortion), more data are suppressed. Consequently, the network lifetime increases.

Physical measurement observed by a sensor node is similar to the measurement of its neighboring nodes. As such, the design discipline of spatial RSA is to suppress the spatial similarity by setting thresholds. Iso-map [5] works by detect isolines which are lines with the same value of measurement. The distance between the isolines (or the resolution of the isolines) depends on given thresholds. Only sensor nodes reside on isolines need to send its readings to the BS.

Joint design on temporal and spatial RSA is possible, but high computation (high complexity) and communication overheads (exchange information) make it infeasible to deployed in a WSN in practice. For example, temporal-spatial RSAs face a dilemma of choosing temporal and spatial RSAs. Meng et al. [6] proposed a temporal-spatial RSA, event contour, to improve network lifetime of WSN. Event contour allows sensor nodes to suppress their reports when both temporal and spatial redundancy suppression conditions are satisfied. However, as following this rule, event contour dose not improve the performance of WSN as compared with temporal and spatial RSAs.

In this paper, we proposed a novel temporal-spatial RSA, Codebook-based Redundancy Suppression Mechanism (CRSM), which utilizes vector quantization technique to construct codebooks based on the distribution of physical phenomena, and use the codebooks to conduct temporal-spatial redundancy suppression. By using CRSM, sensor nodes is able to reduce both transmission rate and the number of

Huan Chen and Bo-Chao Cheng are with National Chung Cheng University, Chiayi, Taiwan (e-mail: huan@ee.ccu.edu.tw, bcheng@ccu.edu.tw).

Chih-Chuan Cheng, Yi-Geng Chen and Yu Ling Chou are with National Chung Cheng University, Chiayi, Taiwan (e-mail: m9288,m9762@cn.ee.ccu.edu.tw lousiechou@billion.com.tw).

Manuscript received June 28, 2009; accepted July 16, 2009.

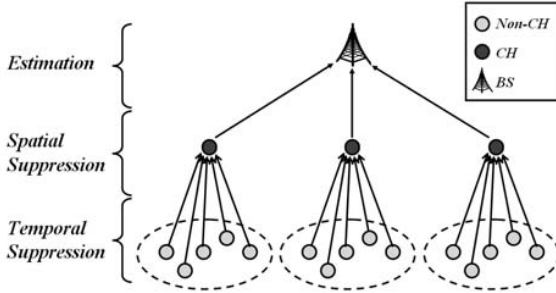


Fig. 1. Cluster-based WSN

transmission, thus improving the network lifetime of WSN. Moreover, the characteristics of codebooks allow us to predict the network lifetime of WSN. Simulation result shows that the performance gain is up to 66% in the best case for CRSM compared to other RSAs.

The remainder of this paper is organized as follows. Section II discusses the background of cluster-based WSN. Section III presents our proposed CRSM, which is followed by network lifetime analysis in the following section. Section V describes simulations to evaluate the performance of CRSM. Finally, Section VI draws a conclusion and outlines possible future extensions of this work.

II. SYSTEM MODEL

A. Cluster-based Routing Protocol

We consider a cluster-based WSN, as shown in Fig. 1, and employ LEACH protocol [3] as a platform to implement our proposed model. In LEACH protocol, the operation time is divided into rounds. Each round consists of two phases: set-up phase and steady state phase. In set-up phase, each sensor node selects itself as a cluster head (CH) at random. After that, each CH broadcasts its status to non-cluster heads (non-CHs). Upon receiving the messages transmitted by the CHs, each non-CH delivers join-request message to the closest CH to join its cluster. Each CH in turn creates a TDMA schedule for its cluster members, and sends the TDMA schedule to the cluster members. In steady state phase, each non-CH sends sensed data to the CH during allocated time slot. And then, each CH aggregates the received data, and transmits the aggregated data to the BS, which will reconstruct the original data. Intuitively, temporal RSA is implemented on non-CHs, and spatial RSA is carried out on CHs. Furthermore, the BS will estimate the suppressed data according to the received data. Note that whatever RSA is used, a given distortion should be guaranteed.

III. CODEBOOK-BASED REDUNDANCY SUPPRESSION MECHANISM (CRSM)

The operations of Codebook-based Redundancy Suppression Mechanism (CRSM) comprises three phases: (OP1) Codebook construction from training dataset, (OP2) Performing codebook-based temporal-spatial suppression for incoming dataset at each sensor node, and (OP3) Performing de-coding and re-construct the data set at the BS. Each phase is described in details as below.

- **(OP1:Codebook Construction)** CRSM designs its codebook based on the statistical information. The design goal is to minimize the coding bit rate R of each sensor, given that the resulting distortion (at this coding rate), $D(R)$ does not exceed the target distortion upper bound constraint \bar{D} . It can be expressed as,

$$\begin{aligned} \min \quad & R \\ \text{s.t.} \quad & D(R) \leq \bar{D}. \end{aligned} \quad (1)$$

In this paper, vector quantization (VQ) [7] is used here as an example. VQ is a well-known entropy coding technique which can meet the distortion requirement while minimizing the coding rate. Given a distortion upper bound \bar{D} , the minimum rate R_{min} can be determined, which satisfies the distortion requirement $D(R_{min}) \leq \bar{D}$. As such, the number of subgroups n can be determined. Given a set of arbitrarily data points $\{x(t)\}$, they are first partitioned into n subgroups, $S=\{s_1, s_2, \dots, s_n\}$. The dynamic data range for each subgroup is called the stepsize. In addition, the value of the centroid point is used as the representative q_i for the i -th subgroup. The representative q_i for the subgroup is also known as the value of the quantization level. It can be proved that if $x(t) \in s_i$, then $|x(t) - q_i| < |x(t) - q_j|, \forall i \neq j$. In fact, q_i is the best estimation for the data in the i -th subgroup if they were not transmitted.

Following the same design disciple of the entropy encoding technique, a codeword c_i is assigned for each q_i . The length of each quantization level q_i is $-\log Pr(q_i)$ and the average transmission rate is the entropy of Q , i.e.,

$$R = H(Q) = - \sum_{i=1}^n Pr(q_i) \log Pr(q_i). \quad (2)$$

By solving Eq. 1, a codebook $C = \{c_1, c_2, \dots, c_n\}$ associated with the target distortion \bar{D} is obtained.

- **(OP2:Codebook-based Temporal-spatial Redundancy Suppression)** Each sensor node computes its own codebook. We assume that probability distribution of physical phenomena throughout sensor field is homogeneous such that codebooks generated by the sensor nodes are identical. Once the codebook is built, the codebook is sent to the BS for future decoding use. Furthermore, the codebook is used as a reference to conduct temporal-spatial redundancy suppression. If the incoming data reading $x(t)$ falls in the subgroup s_i , it is encoded as codeword c_i . After that, each non-CH sends its node id NID and codeword c_i to the CH. Additionally, a threshold $\delta = \bar{D}$ is designed to conduct temporal suppression. As such, the criteria to suppress the current sensed data is when $|x(t) - x_l| < \delta$, where x_l is the quantization level of the previous transmitted codeword.

When each CH receives all reports transmitted by its cluster members, it aggregates the received reports by suppressing spatial redundancy. The operation works as follows. First, each CH merges the received reports into groups, in which the codewords of group members are equal. Second, the overlapped codewords in each group

are suppressed. In other words, only one codeword is sufficient to represent other reports in each group. Finally, each CH sends the aggregated data to the BS.

- **(OP3:Reconstruction)** Since the BS keeps the codebooks for each sensor, as such the BS is able to decode the codeword. After receiving the aggregated data, the BS first maps the codeword in each group to a specific quantization level, say q_i , which is the representative value of the group. Define $\hat{x}(t) = q_i$, which is the estimated value of the sensed data at time t . Next, the BS will reconstruct the missing reports based on the received information. In this paper, the BS assume that the suppressed data remains unchanged, i.e., $\hat{x}(t) = x_l$.

IV. NETWORK LIFETIME ANALYSIS

Because sensor nodes have limited battery, the predictability of network lifetime of WSN facilitates users to manage WSN applications, e.g., biomedical application. Network lifetime is defined as the interval between the instant that WSN starts functioning and the instant that a node in the WSN exhausts its energy. In this paper, CRSM uses the transmission rate R derived in eq. 2 to estimate the network lifetime of WSN. The energy consumption of transmission of sensor nodes is based on first order radio model (as shown in fig. 2. Following the notation and convention as in LEACH [3], the first order radio model is given below:

$$E_{Tx}(L, d) = E_{Tx-elec}(L) + E_{Tx-amp}(L, d) \quad (3)$$

$$= \begin{cases} L \cdot E_{elec} + L \cdot \epsilon_{fs} \cdot d^2 & \text{if } d < d_0 \\ L \cdot E_{elec} + L \cdot \epsilon_{mp} \cdot d^4 & \text{if } d \geq d_0 \end{cases}$$

The total energy consumption $E_{Tx}(L, d)$ for transmitting L -bit message over a distance d can be expressed as a sum of two terms, $E_{Tx-elec}(L)$ and $E_{Tx-amp}(L, d)$. $E_{Tx-elec}$ is the electronics energy consumption, which is related to factors such as the digital coding, modulation, filtering and spreading schemes of the signal. E_{Tx-amp} is the amplifier energy consumption to transmit acceptable BER (bit error rate) for signals transmitted to a receiver. $E_{Tx-elec}$ can be further expressed in terms of energy consumption for a single bit E_{elec} , while E_{Tx-amp} can be further expressed in terms of ϵ_{fs} or ϵ_{mp} , depending on the transmitter amplifier mode that applied. They are power loss factors for free space (d^2 loss) and multi-path fading (d^4 loss), respectively. The threshold d_0 can be determined by equating the two expressions. This radio model reflects the fact of using several immediate short-range transmissions to send data is more energy-efficient than using a single long-range transmission. We will follow the same design rule as well. By equating the two expressions, an empirical value of $d = d_0 = \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}}}$ can be derived.

The total energy consumption $E_{Rx}(L)$ for receiving L -bit message equivalent to $E_{Tx-elec}(L)$. Notice that, the energy consumption for $E_{Rx}(L)$ depends only on how many bits it is receiving rather than the distance d the message is transmitted from, i.e., $E_{Rx}(L, d) = E_{Tx-elec}(L) = L \cdot E_{elec}$.

Assume the size of sensor network is $M \times M$, the number of sensor nodes in the sensor network is N , and the number of CHs is k . At each round, there are approximately $(N/k - 1)$

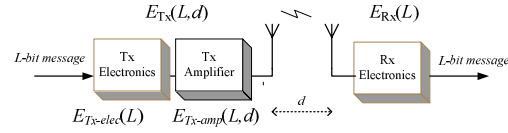


Fig. 2. First Order Radio Model

non-CHs in a cluster. Each non-CH sends its readings with probability $(1 - \rho_T)$, where ρ_T is the probability that the incoming data reading $x(t)$ is within the dynamic range δ of the previous transmission x_l , i.e.,

$$\rho_T = \sum_{i=1}^n Pr(|x(t) - q_i| \leq \delta | x_l \in s_i) Pr(x_l \in s_i). \quad (4)$$

Thus, the energy consumption of a non-CH is calculated by

$$E_{non-CH} = k \cdot E_{Rx}(\text{HDR} + \text{ADV}) \quad (5)$$

$$+ E_{Rx}(\text{HDR} + \text{TDMA} \cdot (1 - \rho_T) \cdot (N/k - 1))$$

$$+ (1 - \rho_T) \cdot E_{Tx}((\text{HDR} + \text{JOIN}), d_{Bcast}^4)$$

$$+ (1 - \rho_T) \cdot E_{Tx}((\text{HDR} + (R + R_{NID})), d_{toCH}^2).$$

where HDR is the length of packet header, ADV is the length of advertisement message broadcasted by CHs, TDMA is the length of TDMA schedule, JOIN is the length of join-request message, R_{NID} is the the length of node id, d_{Bcast} is the distance of broadcast, and d_{toCH} is the distance between non-CH and CH. According to [3], d_{toCH} is derived, i.e., $d_{toCH} = M^2 / (2\pi \cdot k)$. The first term is the energy consumption of receiving advertisement messages transmitted by CHs. The second term is the energy consumption of receiving TDMA schedule. Due to temporal suppression, the length of payload of TDMA schedule equals $\text{TDMA} \cdot (1 - \rho_T) \cdot (N/k - 1)$. The last two terms are the energy consumption of conveying the join-request message and the sensed data.

The total energy consumption for a CH is expressed as

$$E_{CH} = (1 - \rho_T) \cdot (n/k - 1) \cdot E_{Rx}(\text{HDR} + \text{JOIN}) \quad (6)$$

$$+ (1 - \rho_T) \cdot (n/k - 1) \cdot E_{Rx}(\text{HDR} + (R + R_{NID}))$$

$$+ (1 - \rho_T) \cdot ((R + R_{NID})) \cdot E_c \cdot (n/k)$$

$$+ E_{Tx}((\text{HDR} + \text{ADV}), d_{Bcast}^4)$$

$$+ E_{Tx}((\text{HDR} + \text{TDMA} \cdot (1 - \rho_T) \cdot (N/k - 1)), d_{Bcast}^4)$$

$$+ E_{Tx}((\text{HDR} + (n/k) \cdot (1 - \rho_T) \cdot ((\rho_S \cdot R + R_{NID}))), d_{toBS}^4).$$

where d_{toBS} is the distance between the CH and the BS, and ρ_S is the probability that any two codewords among the data received by a CH are the same, i.e.,

$$\rho_S = \sum_{i=1}^n Pr(x_u(t) \in s_i | x_v(t) \in s_i) Pr(x_v(t) \in s_i), \quad \forall u \neq v. \quad (7)$$

where $x_u(t)$ and $x_v(t)$ are the sensed value of the sensor node u and v , respectively.

With deriving the energy consumption of non-CH and CH, the energy dissipated in a cluster during a round is obtained by calculating

$$E_{cluster} = E_{CH} + (n/k - 1) \cdot E_{nonCH} \quad (8)$$

and the total energy for a round is

$$E_{total} = k \cdot E_{cluster}. \quad (9)$$

Hence, when the initial energy of sensor nodes E_i is homogeneous, the network lifetime of WSN LT_{net} can be determined, i.e.,

$$LT_{net} = \frac{N \cdot E_i}{E_{total}} \quad (10)$$

V. NUMERICAL ANALYSIS AND SIMULATE RESULT

Performance evaluation for CRSM is conducted by analysis and simulation. The related parameters are listed as below:

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Transmitter/Receiver Electronics E_{elec}	50 nJ/bit
Aggregation Electronics E_C	50 pJ/bit
Transmit Amplifier ϵ_{fs} ($d < d_0$)	10 pJ/bit/m ²
Transmit Amplifier ϵ_{mp} ($d > d_0$)	0.0013 pJ/bit/m ²
Initial Node Energy E_0	0.2J
Network Size	100m × 100m
Number of Nodes	100 nodes

The packet sizes used in the evaluation are shown in table II.

TABLE II
THE LIST OF PACKET SIZES

Packet	Size
ADV	2bytes
JOIN	16bytes
HDR	20bytes
TDMA	16bytes × cluster members

Assume that physical phenomena is normally distributed with mean 30 and standard deviation 1. The network lifetime is evaluated for CRSM, TEEN, iso-map, and contour map under a variety of distortion constraints (varying from 0.2 to 1). The details of the evaluation are discussed in the following section.

A. Numerical Results

We first determine ρ_T and ρ_S under distortion constraints of 0.2, 0.4, 0.6, 0.8, and 1, respectively. The results are shown in table III.

Intuitively, as maximum acceptable distortion increases, the values of ρ_T and ρ_S increase. Finally, by calculating eq. 10, the network lifetime of WSN is obtained (as shown in table IV).

TABLE III
THE VALUES OF ρ_T AND ρ_S UNDER DISTORTION CONSTRAINTS OF 0.2, 0.4, 0.6, 0.8, AND 1

\bar{D}	ρ_T	ρ_S
0.2	0.1133	0.0942
0.4	0.2255	0.1674
0.6	0.3351	0.2366
0.8	0.4386	0.2738
1	0.536	0.3209

TABLE IV
THE NETWORK LIFETIME EVALUATION FOR CRSM UNDER DISTORTION CONSTRAINTS OF 0.2, 0.4, 0.6, 0.8, AND 1

\bar{D}	LT_{net}
0.2	426
0.4	490
0.6	563
0.8	648
1	755

B. Simulation Results

Our simulation platform is based on NS2, which is a discrete-event-driven simulation platform targeted at networking research. Fig. 3 shows performance for CRSM, TEEN, iso-map, and contour map under high to low distortion constraints, varying from 0.2 to 1. First of all, the numerical result of CRSM is close to the simulation result of CRSM. The difference between these two results is due to the way of cluster head selection. Cluster head selection in the simulation result is at random, while Cluster head selection in numerical result is fixed (the number of CH is always k). Next, fig. 3 shows that CRSM overcomes TEEN, iso-map, and contour map. CRSM outperforms at least 23%, 44%, and 66% of the network lifetime of TEEN, iso-map, and contour map, respectively. Fig. 3 also shows that the network lifetime of TEEN is better than that of iso-map only when distortion constraint is less than 0.8. That is because in iso-map only sensor nodes resided on isolines send their readings to the BS, thereby exhausting their battery rapidly. Finally, as we can see, contour map performs the worst among all RSAs because it rarely suppresses incoming data readings.

VI. CONCLUSION

In this paper, we proposed a novel temporal-spatial redundancy suppression algorithm (RSA), Codebook-base Redundancy Suppression Mechanism (CRSM). CRSM adopts vector quantization to generate a codebook, which is easy to be used to implement temporal-spatial RSA. CRSM not only achieves power saving and reliability on performance in WSN, but also provides the predictability of network lifetime. The performance evaluation is evaluated for CRSM, TEEN, iso-map, and contour map by analysis and simulation. Simulation result shows that the network lifetime of CRSM outperforms at least 23% of that of other RSAs. Our future work targets on the energy efficiency for WSN in which probability distribution of physical phenomena is heterogeneous.

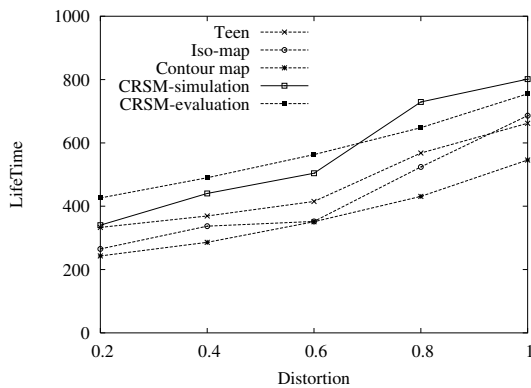


Fig. 3. Performance comparison on energy consumption for CRSM, TEEN, iso-map, and contour map

ACKNOWLEDGMENT

This research was supported in part by the National Science Council (NSC) in Taiwan under the grant numbers NSC-97-2220-E-194-011 and NSC-97-2221-E-194-043.

REFERENCES

- [1] P. Juang et al., Energy-efficient computing for wildlife tracking: design tradeoffs and early experiences with zebnet, In Proceedings of ASPLOS, Oct. 2002.
- [2] A. Mainwaring and J. Polastre and R. Szewczyk and D. Culler and J. Anderson, Wireless sensor networks for habitat monitoring, In Proceedings of WSN, Sep. 2002.
- [3] Wendi B. Heinzelman and Anantha P. Chandrakasan and Hari Balakrishnan, An application-specific protocol architecture for wireless microsensor networks, T-COMM, vol. 1, no. 4, pp. 660-670, Oct. 2002.
- [4] A. Manjeshwar and D. P. Agrawal, TEEN : A protocol for enhanced efficiency in wireless sensor networks, 1st International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing, April 2001.
- [5] Yunhao Liu and Mo Li, Iso-map: Energy-efficient contour mapping in wireless sensor networks, IEEE ICDCS, June 2007.
- [6] X. Meng and L. Li and T. Nandagopal and S. Lu., Contour maps: Monitoring and diagnosis in sensor networks, Computer Networks, 2006.
- [7] I. Koutsopoulos and S. Toupis and L. Tassiulas, On the relation between source and channel coding and sensor network deployment, International Workshop on Wireless Ad-hoc networks (IWWAN) 2005, London UK.



Huan Chen received his B.S. and M.S. (Electrical Engineering) degrees from the National Tsing Hua University (NTHU), Hsing Chu, Taiwan, R.O.C., in 1993 and 1995, respectively. He earned his Ph.D. (Electrical Engineering) degree from the University of Southern California (USC), Los Angeles, in 2002. He also served as the TA and RA in Signal and Image Processing Institute, Integrated Media Systems Center and Electrical Engineering-Systems Department at University of Southern California during his Ph.D. study. He is a member of IEEE since 1999

and he is author of the book "Radio Resource Management for Multimedia QoS Support in Wireless Cellular Networks", Kluwer Academic Publishers, Oct. 2003 (ISBN: 1-4020-7623-1). In addition to his academic career experiences, he also worked for Xitec Inc., Summitec Co. and Intervideo Inc. as a summer intern and software engineer from June 1997 to January 2001. Since February 2002, he has been with the Electrical Engineering Department at National Chung Cheng University, Chiayi (Taiwan). His principle research interests include QoS support, Ad Hoc routing and network optimization. Other research interests include Ubiquitous computing and network security.

Bo-Chao Cheng is an Assistant Professor of Department of Communication Engineering at National Chung-Cheng University. Cheng received a PhD degree in CIS from New Jersey Institute of Technology in 1996. After graduations, he also worked for Transtech Network (2000-2002), Bellcore (1998- 2000) and Racal DataCom (1996-1998) respectively. His broad interests include network security, network management and real-time embedded system design.



Chih-Chuan Cheng received his B.S. in Electrical Engineering from Minghsin University of Science and Technology, Taiwan, in 2000, and his M.S. degree in Electrical Engineering from National Chung Cheng University (NCCU), Taiwan, in 2005. He is now a PhD student of Department of Electrical Engineering at National Chung Cheng University. His research interest includes energy-efficient routing in wireless sensor networks. Other research interests include task planning and resource management in wireless multimedia networks.

Yi-Geng Chen received his B.S. in Computer Science and Information Engineering from I-SHOU University, Taiwan, in 2008. He is now a master student of Department of Electrical Engineering at National Chung Cheng University. His research interest includes energy-efficient routing in wireless sensor networks.

Yu Ling Chou received her B.S. in Department of Applied Mathematics from Tatung University, Taiwan, in 2000. She is now a master student of Master Program of E-Learning for Information and Communications Engineering at National Chung Cheng University. Her research interest includes energy-efficient routing in wireless sensor networks.