Maximizing Lifetime in Clustered WSNs with Energy Harvesting Relay: Profiling and Modeling

Pengfei Zhang*†, Hwee-Pink Tan‡, Gaoxi Xiao§ and Yi Yu¶
*§¶School of Electrical & Electronic Engineering, Nanyang Technological University (Singapore)
†‡Sense and Sense-abilities, Institute for Infocomm Research (Singapore)
Email: {pzhang1@e.ntu.edu.sg, hptan@i2r.a-star.edu.sg, egxxiao@e.ntu.edu.sg and yyu6@e.ntu.edu.sg }

***Abstract*—Inspired by clustering and energy harvesting techniques, we study multiple-cluster wireless sensor networks with energy harvesting (EH) sensors serving as relay for cluster heads. In this paper, we derive the model for realistic energy harvesting rate. Then we propose distributed matching algorithm for EHs to serve as relay for CHs. The proposed algorithm could find optimal/near-optimal CH-EH matching in short time and still achieve good performance. We evaluate the performance of our method through theoretical analysis as well as simulation.**
Keywords: Wireless sensor network, energy harvesting wireless sensor, clustering algorithm, network lifetime

I. Introduction

Nowadays, wireless sensor networks (WSNs) are widely used in many monitoring applications, such as auto-mobile, structure health monitoring, military, health-care etc. Numerous research efforts are carried out by researchers worldwide to improve the performance of WSNs, e.g. [1]–[3].

One of the critical limitations in conventional battery powered WSNs is finite network lifetime since sensor batteries may not be conveniently replaced or recharged. Among the techniques to maximize network lifetime, clustering techniques [4], [5] make use of advanced data aggregation techniques to aggregate data from sensors and forward them to the data sink. A clustered WSN is typically composed of many clusters and a base station (BS), the latter which acts as a data sink. Each cluster comprises a cluster head (CH) and non cluster heads (NCHs). There are typically three phases in clustering protocols for WSNs: *(i) CH selection, (ii) cluster formation and (iii) data transmission*. In most network scenarios, CHs usually strongly affect network lifetime since CHs have to communicate with BS through a longer distance than the distance between NCHs and CHs. Cluster formation also affects the lifetime of CHs since inappropriate cluster formation may force either CH or NCH to be depleted of energy sooner.

An alternative way to overcome the shortage of limited battery capacity is to use energy harvesting (EH) technology [6] to harvest energy from environment. Since the deployments of large-scale WSNs composed solely of EH sensors remain impractical in the near future due to high costs and low achievable duty cycles, deploying EH sensors sparsely in WSNs is typically a more practical approach [7], [8].

Clustering methods and energy-harvesting techniques can be combined together to prolong the network lifetime. Due to fluctuating energy harvesting rates [9], EH sensors may not be suitable to serve as CH nodes that need to operate continuously. In this paper, EH sensors are deployed sparsely and $matched$ with CHs to help them relay data to BS without sensing from environment. To do this effectively, it is necessary to model the fluctuating energy harvesting rate for optimal scheme design. As far as we know, we can model the solar energy harvesting rate as time series values. Autocorrelation is a property of time series value, it is the cross-correlation of a signal with itself. Autocorrelation is most suitable for time series values with high correlated data. Informally, it is the similarities between observations as a function of the time lag between them. Solar energy harvesting rate can be modeled with autocorrelation property due to its high correlation with itself [6]. Moving average property is another important property of solar energy harvesting rate. On the other hand, the EH-CH matching must be done optimally so that the energy consumption of CHs is reduced by communicating through a shorter distance for a certain fraction of time (e.g., when EH nodes are up and working) to relay instead of direct communication with BS.

In our previous work [10], we adopt a distributed approach to solve the above problem, where we considered nodes powered by ambient energy harvesting as relay nodes for CHs, and proposed joint clustering and distributed matching algorithms for network lifetime maximization. In this paper, we extend the study in previous work to deal with fluctuating energy harvesting rate profile instead of constant energy harvesting rate. Our objective is to model the fluctuating energy harvesting rate profile and develop distributed clustering and matching algorithms to optimize the network lifetime.

We organize the rest of papers as follows: In Section II, a brief survey of some closely related work is provided. In Section III, we present our network and assumptions. We also present models to characterize the fluctuating energy harvesting rate and show the simulation results for various models. In Section IV, we propose distributed clustering and CH-EH matching algorithms. Extensive simulation results and discussions are also provided. Finally, Section V concludes the paper and provides directions for future research.

II. Related Work

Amongst the works [11]–[13] that maximize network lifetime, Heinzelman *et al.* proposed LEACH which uses randomized rotation of the cluster head to avoid quickly draining the battery of any sensor in the network [11]. Qing *et al.* proposed and evaluated the distributed energy-efficient clustering scheme (DEEC) for heterogeneous WSNs (i.e., CHs and NCHs have different energy) [12]. A novel energy efficient clustering scheme (EECS) was proposed in [13], which better suits periodical data gathering applications. These works consider battery powered sensor nodes where CHs directly communicate with BS. Our work extends the distributed algorithm by also utilizing EH sensors as relay for CH nodes.

There have also been some studies on modeling of solar energy harvesting rate [14], [15] and clustering in WSNs with EH nodes [7], [16], [17], typically assuming that the network is solely composed of EH sensors which have infinite lifetime. In the area of energy harvesting rate modeling, Kansal et. al used moving average (MA) filter to simulate the future energy profile [14]. Piorno et. al used Weather Conditioned Moving Average (WCMA) algorithm as the solar energy simulation algorithm [15]. Except moving average property, we enhance the previous work by also considering the autocorrelation property of solar harvesting rate. In our work we consider the auto-regressive model including Arima [18] and Garch model [19], which are widely used time series simulation models and consider the autocorrelation property. We compare our model with MA model [14]. In the area of clustering algorithms in energy harvesting WSNs, Islam *et al.* considered a hybrid WSN which comprises both battery-powered and EH nodes [7]. However, they let EH nodes serve as CHs with a higher probability than battery-powered nodes. To the best of our knowledge, only our previous works [10], [20] has studied on schemes maximizing network lifetime where EH nodes serve as relay nodes for CHs.

Numerous matching algorithms have been proposed in optical networks. The representative algorithms include PIM (Parallel Iterative Matching) [21] and iSlip [22] that use random and round-robin approaches respectively for matching between input and output ports. Both algorithms are iterative and proven to converge in $\mathcal{O}(\log N)$ iterations on average. We propose an EH-CH matching algorithms and benchmark it against the PIM approach [21].

III. Solar energy harvesting profiling and characterization

In this section, we first propose our network model (shown in Figure 1) and assumptions. Then we present the experiment setup (shown in Figure 2) to measure the energy harvesting rate from solar panel. Then we present Arima model; followed by the Garch model. We validate our results through simulations in R and Matlab for a 2-D network with data transmission rate of 25 kbps per node. Specifically, we compare the accuracy of different models in Section III-B.

A. Proposed Model

In this paper, we assume a large scale network with N_s sensors all with the same amount of energy initially and N_e EH nodes are randomly deployed in fixed locations. The sensor nodes are partitioned into N_c clusters, each comprising one CH. Each EH serves as relay for one CH, which can be different at different times. Our network scenario is depicted in Figure 1. Table I shows the notations used throughout the paper.

TABLE I
NOTATIONS USED THROUGHOUT THIS PAPER

Notation	Description
N_s	Number of sensors in the network
N_c	Number of clusters in the network
N_e	Number of EH nodes in the network
p	The order of the autoregressive in Arima
d	The order of integrated in Arima
q	The order of moving average in Arima
L	The lag operator
ϕ_i	The parameter of autoregressive in Arima
θ_i	The parameter of moving average part in Arima
ε_t^{arima}	Error term in Arima
s	The order of Garch term σ^2 in Garch
t	The order of ARCH term ϵ^2 in Garch
ε_t^{garch}	The error term in Garch
σ_t	Time-dependent standard deviation in Garch
z_t	White noise random variable
α_i	The parameter of ARCH term
β_i	The parameter of GARCH term
E_{CH_i}	residual energy in CH_i at the beginning of each round
$E^{on}_{CH_i,EH_j}$	residual energy in CH_i after each round with EH_j as relay
P_{CH_i,EH_j}	transmission power between CH_i and EH_j
T	Duration of each round
E_{EH}	energy stored at EH node
E_s	energy stored in battery for each sensor

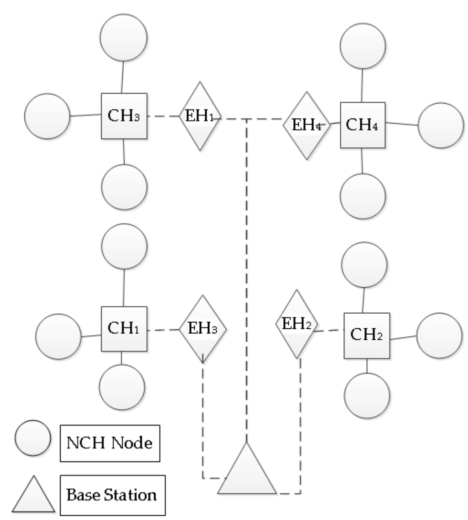

Fig. 1. Abstract model of multiple EH nodes serve as relay for CHs

We assume TDMA and CSMA based communication, as in [11]. The network lifetime is measured in number of TDMA rounds the network can operate until the first node dies. Within each round, each NCH transmits a packet of 2000 bits to its respective CH in a TDMA time slot. Then CHs forward data to BS using CSMA approach, either directly, or via EH node.

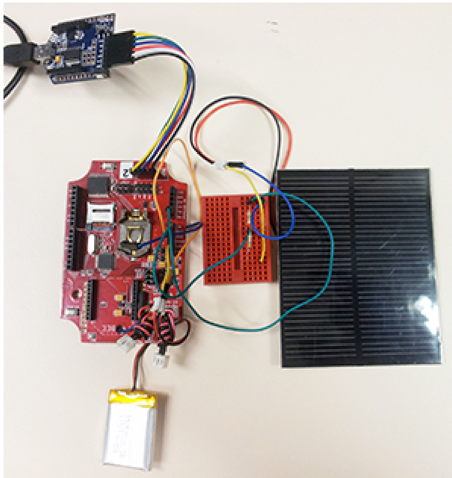

Fig. 2. Experimental setup

We connect the solar panel to Seeeduino board [23] to measure energy harvesting rate from solar panel. The energy harvesting rate is sampled every one second. We use $\widehat{X_t}$ to denote the energy harvesting rate measured from Seeduino board.

The **Arima** model [18] is defined as follows:

$$(1-\sum_{i=1}^{p}\phi_i L^i)(1-L)^d X_t^{arima} = (1+\sum_{i=1}^{q}\theta_i L^i)\varepsilon_t^{arima}$$

where $(1-\sum_{i=1}^{p}\phi_i L^i)X_t^{arima}$ represents the autoregressive part and $(1+\sum_{i=1}^{q}\theta_i L^i)\varepsilon_t^{arima}$ represents the moving average part. X_t^{arima} is the simulated energy harvesting rate in time slot t. The other parameters are shown in Table I. We take the energy harvesting rate $\widehat{X_t}$ measured from Seeduino board as input. Then we find the optimal parameters for Arima model including (p,d,q) and ϕ_i, θ_i such that the simulated series X_t^{arima} is close to $\widehat{X_t}$.

The procedure to adopt **Arima** model is as below:

- Step 1: Take the measured time series value $\widehat{X_t}$ as input.
- Step 2: Find the optimal parameters for (p,d,q) and ϕ_i, θ_i using $auto.arima$ function in R.
- Step 3: Run the model to generate the simulated series X_t^{arima}.

The **Garch** model is defined as below:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{r}\alpha_i {\varepsilon_{t-i}^{garch}}^2 + \sum_{i=1}^{s}\beta_i\sigma_{t-i}^2 \qquad (1)$$

$$\varepsilon_t^{garch} = \sigma_t z_t$$

$$X_t^{garch} = E_x^{garch} + \varepsilon_t^{garch}$$

Similarly as Arima model, Garch model is also composed of two parts. $\sum_{i=1}^{r}\alpha_i {\varepsilon_{t-i}^{garch}}^2$ represents the ARCH part and $\sum_{i=1}^{s}\beta_i\sigma_{t-i}^2$ represents the Garch part. $\varepsilon_{t-i}^{garch}$ is the simulated error term from Garch model. X_t^{garch} is the simulated time series values from Garch model. E_x^{garch} is the mean value derived from the best fitted Garch model. α_0 is a constant given by the best fitted Garch model. Similarly as Arima model, the procedure to adopt **Garch** model is as below:

- Step 1: Take the measured time series value $\widehat{X_t}$ as input.
- Step 2: Find the optimal parameters for (t,s) and α_i, β_i using $garchFit$ function in R.
- Step 3: Run the model to generate the simulated series X_t^{garch}.

B. Experiment data

We have the experiment data shown in Figure 3. There are five data sets measured at different time. Each data set lasts for a duration of around twenty minutes. Due to the page constraint, we use our simulation algorithms to model data set 4 and 5. Note that the models are also applicable for the remaining data sets.

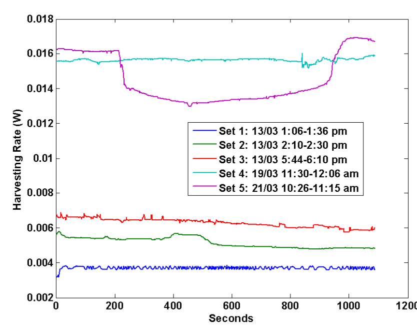

Fig. 3. Experimental data

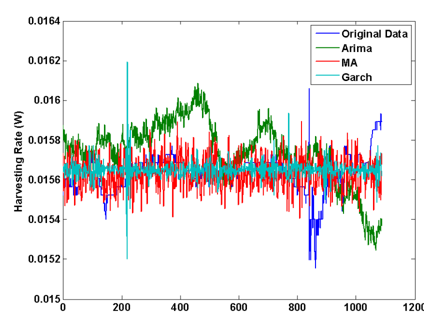

Fig. 4. Comparison of simulation between Arima, MA and Garch in data set 4

We compare Arima(1,1,1) ($\phi_1 = 0.1180$, $\theta_1 = -0.6089$) and Garch(1,1)($\alpha_1 = 0.051$, $\beta_1 = 0.0494$) with moving average model (MA(1)) [14] for data set 4. The simulation

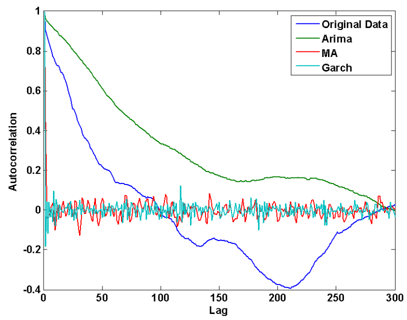

Fig. 5. Comparison of autocorrelation between Arima, MA and Garch in data set 4

curves by different models are shown in Figure 4. We also plot the autocorrelation curves by different models in Figure 5. The optimality of a model is indicated by its autocorrelation function compared with original data set. The autocorrelation curve from Arima(1,1,1) is most closest to the autocorrelation curve from original data. Garch(1,1) and MA(1) produce non-correlative data, which can be seen in Figure 4. This is because Arima has auto-regressive term in its formulation. Garch can produce data with less deviation compared with MA.

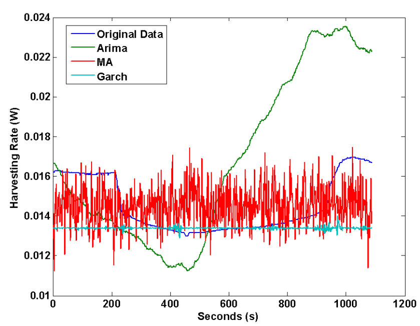

Fig. 6. Comparison of simulation between Arima, MA and Garch in data set 5

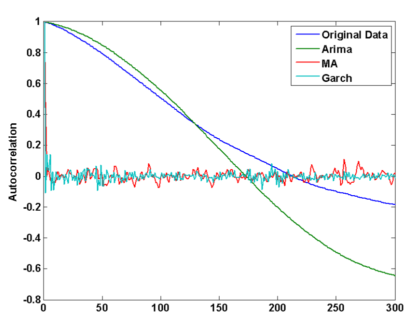

Fig. 7. Comparison of autocorrelation between Arima, MA and Garch in data set 5

The results for data set 5 are shown in Figure 6 and Figure 7, We compare Arima(0,2,1) ($\theta_1 = -0.8844$), Garch (1,1) ($\alpha_1 = 0.593$, $\beta_1 = 0.41$) and MA(1). Similarly as data set 4, Arima returns better auto-correlative data compared with Garch and MA.

IV. Impact on Clustering Algorithm

In this section, we extend the work in [10] by considering fluctuating energy harvesting rate. We adopt the same cluster head selection and cluster formation algorithms. We extend the CH-EH matching algorithm for fluctuating energy harvesting rate in Section IV-A. Then we show the effect of fluctuating harvesting rates on the performance of distributed algorithm in Section IV-B.

A. Distributed CH& EH Matching

We assume EH node can serve as relay for CHs only if its residual energy can sustain transmission of at least one round with duration T. We have:

$$E_{CH_i} - E^{on}_{CH_i,EH_j} = TP_{CH_i,EH_j}$$

$\Rightarrow$

$$E^{on}_{CH_i,EH_j} = E_{CH_i} - TP_{CH_i,EH_j} \tag{2}$$

We enhanced Algorithm 1([10]) by considering the fluctuating energy harvesting rate derived from the models proposed in section III.

B. Performance of Distributed Matching Algorithm

We show the performance of our distributed matching algorithm in Section IV-B. We evaluate the optimality of our approach using Algorithm 1 compared with CC (Centralized clustering [20]) and PIM approach (shown in Section II). In PIM approach, EH node re-selects CH randomly in the setup phase. In CC, CH selection and cluster formation are done in centralized manner. We assume that $E_s = 0.5J$ for all nodes including EH node. Similar to [11], we consider the Friss free space propagation model.

We consider two square regions of network deployment, where the coordinates of the vertices are as follows: Area (I) (100, 100), (100, 200), (200, 100) and (200, 200); Area (II) (0, 0), (0, 200), (200, 0) and (200, 200). CHs and EHs are randomly deployed in these two regions. We assume number of sensors in the region is from $\{100, 125, 150, 175, 200\}$.

The effect of p_{s_i} (this is the probability of sensors to serve as CH in [10]) on lifetime for Case I is shown in Figure 8. We use data set 5 as the fluctuating energy harvesting rate in the simulation. From this figure, we find the optimal p_{s_i} is 0.04, which is adopted in our following simulations.

The comparison between our algorithm (DA) with CC and PIM is shown in Figure 9 for Case I and Figure 10 for Case II. We use data set 5 in Figure 3 as our energy harvesting rate. The result shows that our algorithm outperforms PIM. Specifically, when $N_s = 100$ in Case I, lifetime is 458.35 for DA and 378.6 for PIM. This is because our algorithm selects the EH node to serve the best CH it could find; while PIM considers using EH nodes to randomly serve each CH. Thus, our algorithm maximizes the lifetime until the first node dies

Algorithm 1 Distributed matching algorithm at setup phase

Require: N_c CHs, N_e EHs, $\widehat{X_t}$, X_t^{arima} and X_t^{garch}
Ensure: Matching between CHs and EHs

Assume that the simulated harvesting rate is $\hat{P}_{EH,h}$, which is from $\widehat{X_t}$, X_t^{arima} and X_t^{garch}. At beginning, EH node broadcasts *awake* to CHs in their broadcast range if $E_{EH} + \hat{P}_{EH,h}T > E_s$. Each CH estimates P_{CH_i,EH_j} according to the signal strength and calculates $E^{on}_{CH_i,EH_j}$ according to (2). It also records the smallest $E_i = \min E^{on}_{CH_i,EH_j}$.

while there is unmatched CH **do**

For each unmatched CH:

if there are unmatched EH nodes **then**

It selects the *closest* unmatched EH node and send *request* message including E_i to this EH node.

else[CH node does not receive any *awake* message]

It directly communicates with BS.

end if

For each unmatched EH:

EH node receives several *request* messages from CH nodes in its range. Denote the number of *requests* as N_r, N_r may be equal to 0

if $N_r > 0$ **then**

EH node sends $grant$ to the CH node with smallest E_i.

else[$N_r = 0$]

EH node waits for the next iteration

end if

For each unmatched CH:

CH node may or may not receive *grant* messages from EH nodes in its range, let N_g be the number of *grant* messages it receives.

if $N_g = 1$ **then**

It selects the closest EH node and sends *accept* message to it.

else$N_g = 0$

Return to while loop.

end if

end while

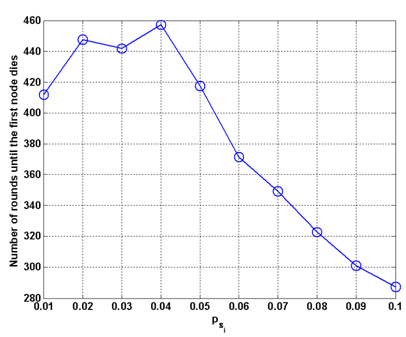

Fig. 8. Effect of p_{s_i} on network lifetime in Case I ($N_e = 4$, $p_{s_i} = 0.04$)

while PIM is trying to find a matching within few iterations.

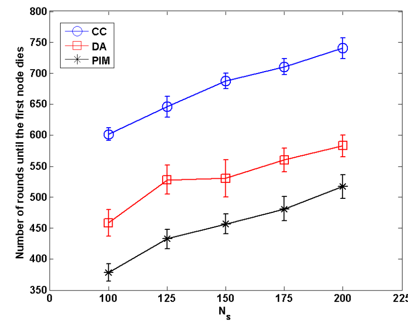

Fig. 9. Comparison between our algorithm with CC and PIM in Case I ($N_e = 4$, $p_{s_i} = 0.04$)

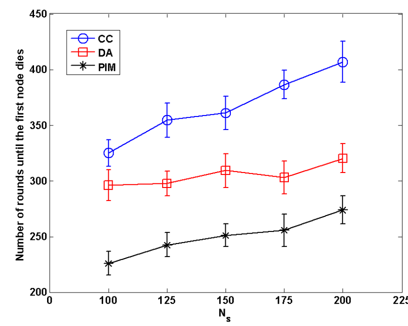

Fig. 10. Comparison between our algorithm with CC and PIM in Case II ($N_e = 4$, $N_s = 100$)

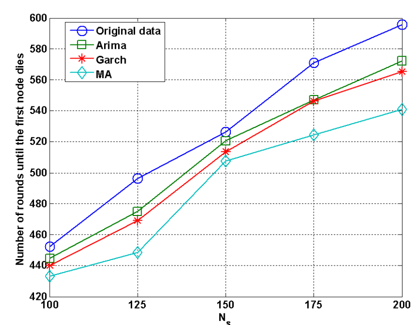

Fig. 11. Comparison between various modeling methods in Case I

DA achieves shorter lifetime compared with CC. Specifically, when $N_s = 100$ in Case I, lifetime is 601.6 for CC. This is because DA uses local information for decision while CC uses global information for optimality. In addition, network lifetime is increasing when N_s is increasing. The reason is that when number of sensors increases, more sensors will be able to serve as CHs, which reduces the chances CHs become *bottleneck* nodes. The conclusion also holds for Case II.

Similar result is shown in Figure 10 for Case II. When $N_s = 100$, network lifetime of DA is 296.1 rounds compared with 222.5 rounds for PIM and 325.05 rounds for CC. The network lifetime for Case II is smaller than Case I because sensors are more sparsely deployed, resulting in higher power consumption of NCHs.

We then compare the performance of Arima, Garch and MA model as shown in Figure 11. Firstly, Arima and Garch models can achieve quite close results compared with original data while MA has the worst performance. Secondly, we may use the difference to measure the accuracy of the models, such as square errors etc. The less the error is, the better the accuracy is. We find the Arima can achieve the lowest MSE, with 40.73 which is smaller compared with Garch and MA, with the MSE of 50.87 and 90.50.

V. Conclusion and Future Work

In this paper, we considered clustered wireless sensor networks (WSN) where CHs either aggregate and forward data directly to BS, or via relay nodes with energy harvesting (EH) capabilities. We first characterize the performance of energy harvesting rate using sensor boards and use Arima and Garch models to fit the harvesting rate profile. Then we proposed efficient distributed matching algorithm between CHs and EHs to maximize network lifetime, where the network lifetime is the duration until the first node runs out of energy. Through theoretical analysis and extensive simulations, we validated the performance of the proposed algorithms.

For future work, we plan to (i) study other configurations of introducing the EH nodes to the network; and (ii) extend our simulations to more realistic models, and implement and evaluate our algorithms in an actual WSN testbed.

Acknowledgement

This work was carried out while Zhang was a postgraduate student at Nanyang Technological University sponsored by A*STAR Graduate Scholarship (AGS) from September 2010.

References

[1] R. Vullers, R. Schaijk, H. Visser, J. Penders, and C. Hoof, "Energy harvesting for autonomous wireless sensor networks," *IEEE Solid-State Circuits Magazine*, vol. 2, no. 2, pp. 29–38, 2010.

[2] T. Arampatzis, J. Lygeros, and S. Manesis, "A survey of applications of wireless sensors and wireless sensor networks," *Proceedings of the 2005 IEEE International Symposium on Intelligent Control*, pp. 719 – 724, june 2005.

[3] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393 – 422, 2002.

[4] A. A. Abbasi and M. Younis, "A survey on clustering algorithms for wireless sensor networks," *Computer Communications*, vol. 30, no. 14-15, pp. 2826 – 2841, 2007.

[5] D. Dechene, A. E. Jardali, M. Luccini, and A. Sauer, "A survey of clustering algorithms for wireless sensor networks," *Computer Communications archive*, Oct 2007.

[6] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: Survey and implications," *IEEE Communications Surveys Tutorials*, no. 99, pp. 1 –19, 2010.

[7] M. Islam, M. Islam, and M. Islam, "A-sleach: An advanced solar aware leach protocol for energy efficient routing in wireless sensor networks," *Sixth International Conference on Networking*, p. 4, april 2007.

[8] B. Medepally and N. Mehta, "Voluntary energy harvesting relays and selection in cooperative wireless networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3543 –3553, november 2010.

[9] C. Bergonzini, D. Brunelli, and L. Benini, "Algorithms for harvested energy prediction in batteryless wireless sensor networks," *3rd International Workshop on Advances in sensors and Interfaces*, pp. 144 –149, june 2009.

[10] P. Zhang, H.-P. Tan, and G. Xiao, "Distributed algorithms for maximizing lifetime in clustered wireless sensor networks using energy-harvesting relay nodes," 2014. [Online]. Available: http://sns.i2r.a-star.edu.sg/wp-content/uploads/2014/12/arxiv2.pdf

[11] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, vol. 2, p. 223, jan. 2000.

[12] L. Qing, Q. Zhu, and M. Wang, "Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks," *Computer Communications*, vol. 29, no. 12, pp. 2230 – 2237, 2006.

[13] M. Ye, C. Li, G. Chen, and J. Wu, "Eecs: an energy efficient clustering scheme in wireless sensor networks," *24th IEEE International Performance, Computing, and Communications Conference*, pp. 535 – 540, april 2005.

[14] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava, "Power management in energy harvesting sensor networks," *ACM Trans. Embed. Comput. Syst.*, vol. 6, no. 4, Sep. 2007.

[15] J. Piorno, C. Bergonzini, D. Atienza, and T. Rosing, "Prediction and management in energy harvested wireless sensor nodes," *Wireless VITAE 2009.*, pp. 6 –10, may 2009.

[16] T. Voigt, A. Dunkels, J. Alonso, H. Ritter, and J. Schiller, "Solar-aware clustering in wireless sensor networks," *Ninth International Symposium on Computers and Communications*, vol. 1, pp. 238 – 243, july 2004.

[17] K. Kinoshita, T. Okazaki, H. Tode, and K. Murakami, "A data gathering scheme for environmental energy-based wireless sensor networks," *5th IEEE Consumer Communications and Networking Conference*, pp. 719 –723, jan. 2008.

[18] T. C. Mills, "Time series techniques for economists," *Cambridge University Press*, 1990.

[19] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, pp. 307–327, 1986.

[20] P. Zhang, G. Xiao, and H.-P. Tan, "Clustering algorithms for maximizing the lifetime of wireless sensor networks with energy-harvesting sensors," *Computer Networks*, Available online 4 July 2013.

[21] T. E. Anderson, S. S. Owicki, J. B. Saxe, and C. P. Thacker, "High-speed switch scheduling for local-area networks," *ACM Trans. Comput. Syst.*, vol. 11, no. 4, pp. 319–352, Nov. 1993.

[22] N. McKeown, "The islip scheduling algorithm for input-queued switches," *Networking, IEEE/ACM Transactions on*, vol. 7, no. 2, pp. 188 –201, apr 1999.

[23] "Seeeduino ADK Main Board - Wiki," http://www.seeedstudio.com/wiki/Seeeduino_ADK_Main_Board/, 2013, [Online; accessed 13-Sep-2013].