Optimal Task-Specific Clustering Scheme in Unreliable Wireless Sensor Networks

Yan Jin[†], *Ling Wang*[†], *Yoohwan Kim*, [††] *and Xiaozong Yang*[†],

[+]Harbin Institute of Technology, Department of Computer Science and Technology, China
[++]University of Nevada, School of Computer Science, Las Vegas, USA

Summary
Wireless sensor networks can be used to collect surrounding data by multi-hop. As diversity of application requirement (energy-efficient centric or latency centric), energy efficiency and data latency are both important designing issues in designing a clustering scheme optimally. In this paper, the energy consumption of two basic clustering schemes is modeled and the impact of unreliable wireless link on sensor networks configuration is discussed. Then an alternate-hop clustering scheme is proposed, which is proven to be more energy-efficient theoretically. Further, the latency within a cluster is analyzed. At last we propose an optimal task-specific clustering scheme in unreliable wireless sensor networks.

Key words:
Wireless Sensor networks (WSN), clustering, reliability, task-specific, energy efficiency, latency

Introduction

A wireless sensor network (WSN) is a type of specialized wireless networks whose purpose is to gather the sensed data from surrounding environment [1, 2]. The nodes in wireless sensor networks are untethered and unattended. There are many applications of WSN, such as environmental monitoring, battlefield tracking and disaster recovery operation [1, 3], a building control system, and smart entertainment devices that adjust audio and video quality based on their surroundings, etc.
The sensors send such collected data, usually via radio transmitters, to a command center (sink) either directly or through a data concentration center (a gateway). Since most of these devices have limited battery life and it is infeasible to replenish energy via replacing batteries on up to tens of thousands of sensors in most of the applications, it is well accepted that a sensor network should be deployed optimally in order to cater to the different application requirement, such as prolong the network lifetime or reduce the data latency. Although designing the energy-efficient ideas throughout the protocol stacks is known to us, there are further literatures focused on MAC schemes (protocols) [4, 5] and network routing schemes (protocols) [6, 7, 8, 9] for energy saving. Comparably, the latency issue has seldom been discussed.
In [4] and [5], authors can save the power consumption by decreasing the time about transmitter turning on and assigning the schedule of slot soundly to avoid signal collision and data retransmission. The literatures of [6] and [7] belong to the flat routing scheme, which maybe well if the density of the network is small or the quantity of delivered data is small. In addition, Rodoplu et al. [8] gave a distributed algorithm called MECN to find a minimum energy cost path from a source node to any destination node. In [9] the first energy-efficient clustering routing protocol LEACH is proposed for sensor networks, and it has been proved to adapt to the large-scale networks and be scalable for future application such as multimedia data transmissions. This approach can group a number of nodes, usually within a geographic neighborhood, to form a cluster. By this way sensors can be managed locally by a cluster head (CH), a node elected to manage the cluster and be responsible for communication between the cluster and the sink node. Simulation result demonstrates that the clustering scheme offers a more flexible balance among reliability, redundancy, and scalability of WSN.
Since then, there are lots of routing schemes derived from [9], such as PEGASIS and Hierarchical-PEGASIS 、 TEEN and APTEEN etc [10]. Although the PEGASIS approaches avoid the clustering overhead of LEACH, they still require dynamic topology adjustment since sensor's energy is not tracked. And TEEN and APTEEN are designed to satisfy critical data collection by adding threshold function. The main drawbacks of the two approaches are the overhead and complexity of implementing threshold-based functions and dealing with attribute-based naming of queries. Recently, Seema et al. [11] give the selected optimal probability to be a CH when the sink node is located in the center of the network and the regular node is far away from its CH at most *K*-hop. However, their results could be better convinced if they consider the overhead of receiving energy, which is an important factor on designing the energy-efficient algorithm [12]. Moreover, because of unreliable wireless link, a formulation based solely on the energy spent in a

Manuscript received July 5, 2006.
Manuscript revised July 25 , 2006.

single transmission is misleading—the proper metric should include the total energy in reliably delivering the packet to its final destination.

To our best knowledge, there is less literatures considering about the influence of unreliable wireless link on the optimal clustering scheme for different applications (energy-centric or latency-centric). In fact, it is an important problem we have to solve.

Therefore, in this paper we focus on energy consumption and latency issues of different clustering schemes in unreliable sensor networks. Consequently, the optimal clustering scheme is given for different applications.

The construction of this paper is organized as follows: in next Section we introduce two basic clustering models. Integrating both transmission energy and receiving energy we derive the energy consumption of unreliable clustering models in Section 3. Besides, an ***A**lternate-**H**op **C**lustering **S**cheme* (AHCS) is proposed to reduce the energy consumption further. The related latency issues of these two basic models are given in Section 4. Following that, we carry out the extensive simulations in Section 5 to validate the derived results. Finally we conclude the paper in Section 6.

2. Preliminaries

Some main notations are given as follows:

R: The radius of a cluster

n: The path loss exponent which equals to 2 or 4

r: The data transmission rate (bit/second)

α: The radio power to run the radio circuitry

β: The transmit amplifier to achieve an acceptable SNR.

Assumptions:

- Nodes are randomly distributed as a two-dimensional Poisson point process with intensity λ. That is, the probability of finding m nodes in a region of area A is equal to $(\lambda A)^m \exp(-\lambda A)/m!$;
- There are two types of nodes: type 0 nodes and type 1 nodes. The type 0 nodes are the (regular) sensor nodes that perform the job of sensing and sending the sensed data to the cluster heads (CHs). The type 1 nodes serve as CHs. The function of CHs is to collecting data from all sensor nodes within a cluster periodically. They are provided with sufficient battery energy. Thus there is no need for a CH election protocol, since the CHs are predetermined. This is called heterogeneous sensor network, which is widely used in [13];
- In practical circumstance, each wireless link has an independent packet error rate of p_{link};

The *energy model* is adopted from [9]:

$$P_{send}(n_1,n_2)=(\alpha+\beta d(n_1,n_2)^n)r \quad (1)$$

$$P_{receive}=\alpha r \quad (2)$$

Where $P_{send}(n_1, n_2)$ is the power consumed by node n_1 when it is transmitting to node n_2 per second, $P_{receive}$ is the power consumed by node n_2 per second, $d(n_1, n_2)$ is the distance between two nodes n_1 and n_2, n is the path loss exponent. r is the data rate for transmission. Here α, β are constant radio parameters, typical values are α=50nJ/bit, β=100pJ/bit/m^2 (n=2) or 0.001pJ/bit/m^4 (n=4) [9].

As part of the analysis, two basic different clustering models are considered:

- *Single-Hop Clustering Scheme* (SHCS): where a regular node transmits a packet to the corresponding CH directly. The model of SHCS is shown in Fig. 1(a);
- *K-Hop Clustering Scheme* (KHCS): where a regular node transmits a packet to the corresponding CH by at most *K* (*K*>1) hops. The model of KHCS is shown in Fig. 1 (b).

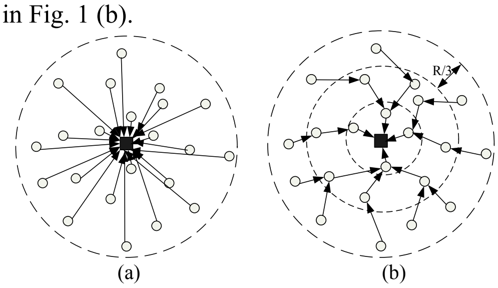


Fig. 1 Structure of: (a) SHCS (b) KHCS (*K*=3).

Obviously, SHCS is the special case of KHCS when *K*=1. In the case of KHCS, we assume the circle of a cluster is divided into concentric rings of thickness *R/K*. That is to say, the transmission range of each node is *R/K*. We note that with a multi-hop communication, if a packet is generated in the l^{th} ring ($1 \le l \le K$), during its journey to the CH, the packet has to travel through each of the inner rings.

Definition 1: The *lifetime* of a cluster is defined as the persistent time until the first sensor node drains out of its initial energy, such node is called *critical node*.

Obviously, the lifetime of a cluster is related to the energy consumption of critical nodes. The relationship between them is inversely proportional.

Definition 2: Define the *communication graph* $G(V,E)$ to be the communication graph of a set of sensors, where each sensor in the set is represented by a node in V, and for any node u and v in V, $E=\{(u,v) \mid u \in V, v \in V, d(u,v) \le r_c\}$, $d(u,v)$ is the Euclidean distance between node u and node v and r_c is the communication range of a sensor node. $N(u)$ is introduced , which presents the neighbor set of node u: $N(u)=\{v \mid v \in V \text{ and } (u,v) \in E\}$.

Definition 3: The *communication graph* $G(V,E)$ *is connected* if and only if a *path* consisting of consecutive edges in E exists between any two nodes u and v.
Next, how the error rate of individual links affects energy consumption and data latency needed to ensure reliable packet delivery within a cluster will be considered.

3. Energy Analysis for Clustering Models

3.1 Reliable Consideration on SHCS

Now we consider a formed cluster. The regular sensors join the cluster of the CH that is closest in terms of communication energy required to reach it.

Proposition 1: In SHCS, given the radius of a cluster R, data rate r and packet error rate p_{link}. When the unreliable transmission is taken into account, i.e., the packets sent by all regular nodes within a cluster were reliable received by a CH, the expected maximal energy consumption of a regular node per second is:

$$E_S = \frac{\alpha + \beta R^n}{1 - p_{link}} r \qquad (3)$$

Proof: First the reliable consideration is ignored. Obviously, under the case of SHCS the regular nodes where locate on the margin of a cluster consume more energy than other nodes in that these marginal nodes have to transmit longer distances from the CH (according to the Eq. (1)). Thus these nodes are called *critical nodes* in SHCS. From Eq. (1), the maximal energy consumption of a critical node is: $E_S^{'} = (\alpha + \beta R^n)r$

Then we will consider the number of transmission (including retransmission) necessary to ensure the reliable (successful) transmission of a packet between a regular node and a CH. This number is a geometrically distribution random variable X, such that:
$Pro\{X=k\}=p_{link}^{\ k-1}(1-p_{link}),\ k=1,2,\ldots$
Therefore, the mean number of individual packet transmission for the successful transfer of a single packet is thus $1/(1-p_{link})$. As a result, the expected maximal energy consumption of a regular node per second is given by (3). ■
From *Proposition 1*, in order to execute T seconds continually, the initial power configuration of every sensor node will be at least TE_S.

3.2 Reliable Consideration on KHCS

According to the feature mentioned in previous Section, the influence of reliability is analyzed as follows:

Proposition 2: In KHCS (K>1), the expected energy consumption of a regular node is maximal when it locates within the 1st ring.
Proof: For each data gathering cycle, we consider the average energy expenditure of a regular sensor node in the l^{th} ring, where l varies from 1 to K. $L(l)$ denotes the average number of packets that a typical node in the l^{th} ring has to relay, then

$$L(l) = \frac{\lambda(\pi R^2 - \pi(lR/K)^2)}{\lambda(\pi(lR/K)^2 - \pi((l-1)R/K)^2)} = \frac{K^2 - l^2}{2l - 1}.$$

Obviously, $L(l)$ is maximal when l=1. It means the nodes of outer ring forward data packages less than those of inner ring. In other words, in KHCS, the lifetime of a cluster depends on the nodes within 1st ring inner. ■

Proposition 3: In KHCS, given the radius of a cluster R, data rate r and packet error rate p_{link}. When the unreliable transmission is taken into account, the expected maximal energy consumption of a regular node per second is:

$$E_K = \alpha(K^2 - 1)r + \frac{1}{1 - p_{link}}[\alpha + \beta(\frac{R}{K})^n]K^2 r \qquad (4)$$

Proof: Based on *Proposition 2*, energy consumption is maximal when l is 1, thus $L(1)=K^2-1$. In this case, the expected energy consumption for these critical nodes is as Eq. (4). The first term of Eq. (4) denotes the energy consumption of receiving K^2-1 packets from other rings averagely; the second term of Eq. (4) denotes the energy consumption of transmitting the packets which come from outer sensor nodes and its own, which needs retransmit $1/(1-p_{link})$ times for reliable consideration. ■
From *Proposition 3*, in order to execute T seconds continually, the initial power configuration of every sensor node will be at least TE_K.

Lemma 1: The *necessary condition* for keeping *network connected* is that for any node u, $|N(u)|\geq 1$, i.e., it has at least one neighbor node in $G(V,E)$. $G(V,E)$ is defined in *Definition 2*.
Proof: We prove this *lemma* by contradiction. Suppose there exists a node v such that it hasn't neighbor node in $G(V,E)$. Obviously, node v is disconnected from $G(V,E)$. Thus $G(V,E)$ isn't a connected graph, which contradicts our assumption. ■
In KHCS, communication graph must be connected so as to ensure a CH can receive the packet from any a sensor node within a cluster. So *Theorem 1* is given:

Theorem 1: Given the intensity λ of Poisson point process and the radius R of a cluster. In order to ensure that the confidential interval of network connectivity is at least 1-ε, the communicating radio range r_c should satisfies: $r_c \geq \hat{r_c}$ where $\hat{r_c} = \sqrt{\frac{1}{\lambda\pi}\log(\frac{\lambda\pi R^2}{\varepsilon})}$

Proof: Given λ and R we know there are $\lambda\pi R^2$ nodes totally. Some symbols are introduced for proof.
A: the event that every node has at least one neighbor node;
A_i: the event that node i has no neighbor node in $G(V,E)$, i.e., $|N(i)|=0$.
To have network connectivity with a probability of at least 1-ε, we have: P(the network is connected) $\geq 1-\varepsilon$. Based on *Lemma 4* we have P(the network is connected)$\geq P(A)$. Therefore, if $P(A) \geq 1-\varepsilon$ then the *Theorem 1* can be satisfied. Thus it should satisfy: $P(\overline{A}) \leq \varepsilon$ where $\overline{A}$ is the supplement of A and $\overline{A} = \bigcup_{j=1}^{\lambda\pi R^2} A_i$. According to the independent property of Poisson point process, $P(\overline{A}) \leq \varepsilon$ => $\lambda\pi R^2 P(A_1) \leq \varepsilon$. And $P(|N(u)|=n)= \frac{e^{-\lambda\pi r_c^2}(\lambda\pi r_c^2)^n}{n!}$ => $P(A_1)= e^{-\lambda\pi r_c^2}$. Therefore, *Theorem 1* can be proven. ■

By differentiating Eq. (4) and equating the result to 0 for minimizing E_K we can derive the optimum K (denoted by K_{opt}^E) easily:

$$K_{opt}^E = R(\frac{\beta(n-2)}{2\alpha(2-p_{link})})^{1/n} \ (n>2) \qquad (5)$$

The value of K_{opt}^E thus obtained depends on the radio parameters (α, β), the radius of a cluster (R), the packet error rate (p_{link}) and path loss exponent (n), regardless of the number of sensor nodes. It can be shown that the second derivative of Eq. (4) is always positive. Hence K_{opt}^E is globally optimal in Eq. (4).

When n=2, E_K-E_S= $\alpha(K^2-1)\frac{2-p_{link}}{1-p_{link}} \geq 0$, which means SHCS consumes energy no greater than KHCS.
Based on *Theorem 1*, since Eq. (4) is a convex function, so if $R/K_{opt} < \hat{r_c}$ K_{opt} is reset as $R/\hat{r_c}$ to keep network connected.

3.3 Reliable Consideration on Alternate-hop Clustering Scheme (AHCS)

In previous sub-sections the maximal energy consumptions of SHCS and KHCS per second are given respectively. In SHCS, the critical nodes locate on the margin of a cluster. In KHCS, the critical nodes locate within 1st ring. So we consider an ***a***lternate-***h***op ***c***lustering ***s***cheme (AHCS) combining SHCS and KHCS to balance the energy consumption of all sensor nodes within a cluster.
Now a sensor node that is located in l^{th} ring ($1 \leq l \leq K$) is considered. For every second, supposing a sensor node works with SHCS φ second and it works with KHCS $1-\varphi$ second alternatively ($0 \leq \varphi \leq 1$). Therefore, the energy consumption of this node per second is:

$$E(l,\phi) = \phi E_S(l) + (1-\phi)E_K(l) \qquad (6)$$

Here $E_S(l) = [\alpha + \beta(\frac{R}{K}l)^n]\frac{1}{1-p_{link}}r$ and

$$E_K(l) = \alpha(\frac{K^2-l^2}{2l-1})r + \frac{1}{1-p_{link}}[\alpha + \beta(\frac{R}{K})^n]\frac{K^2-(l-1)^2}{2l-1}r$$

It can be proven that Eq. (6) is a convex function. It implies that Eq. (6) has a maximal value when l is 1 or K. Then $E_{\max}(l,\phi) = \max\{E(1,\phi), E(K,\phi)\}$. We introduce three variables m_1, m_2 and m_3 for ease of notation:
$m_1=E_S(1)$, $m_2= E_S(K)$ and $m_3= E_K(1)$. So

$$E_{\max}(l,\phi) = \{\phi(m_1-m_3)+m_3, \phi(m_2-m_1)+m_1\}$$

Because $m_1 \leq m_3$ and $m_1 \leq m_2$, $E_{\max}(l,\phi)$ is minimized when $\phi(m_1-m_3)+m_3 = \phi(m_2-m_1)+m_1$. As a result, when $\phi = \frac{m_3-m_1}{m_2+m_3-2m_1}$ the minimized value of $E_{\max}(l,\phi)$ is $E_{\max}^{\min}(l,\phi) = \frac{m_2m_3-m_1^2}{m_2+m_3-2m_1}$. Moreover, we note the energy consumption of SHCS and KHCS is m_2 and m_3 respectively. Obviously, $m_2 > E_{\max}^{\min}(l,\phi)$ and $m_3 > E_{\max}^{\min}(l,\phi)$, which demonstrates our alternate scheme AHCS outperforms both SHCS and KHCS. This observation is verified in Section 5 by simulation.

4. Latency Analysis for Clustering Models

TDMA is used to schedule regular nodes within a cluster. We define a slot in a MAC data frame as 1 unit time.

- SHCS—The mean number of transmission is $1/(1-p_{link})$ for each node, which has mentioned in *Proposition 1*. For each round, the CH should receive the data from other sensor nodes orderly. Therefore, the total latency of a cluster is:

$$T_S^L = \lambda\pi R^2 \frac{1}{1-p_{link}} \qquad (7)$$

- KHCS—Given the value of K, the nodes in l^{th} ring will transmit packets to $(l\text{-}1)^{th}$ ring until they finish receiving the packets from $(l+1)^{th}$ ring ($1 \leq l \leq K$-1). The number of nodes within $(l+1)^{th}$ ring is

$\lambda\pi(\frac{R}{K})^2[(l+1)^2-l^2]$; the number of nodes within l^{th} ring is $\lambda\pi(\frac{R}{K})^2[l^2-(l-1)^2]$. Then we know a node of l^{th} ring receives packets from $\frac{2l+1}{2l-1}$ neighborhood nodes of $(l+1)^{th}$ ring. Therefore, the total latency of a cluster is:

$$T_M^L(K)=\frac{1}{1-p_{link}}\sum_{l=1}^{K-1}\frac{2l+1}{2l-1}+\frac{1}{1-p_{link}}\lambda\pi(\frac{R}{K})^2 \quad (8)$$

Note that the first term of Eq. (8) is the total latency for packets from K^{th} ring to 1^{st} ring, and the second term of Eq. (8) is the latency for packets from 1^{st} ring to the CH.

Theorem 2: In KHCS, given R and λ. The latency of KHCS is minimized if and only if $K_{opt}^L=(2\lambda\pi R^2)^{1/3}$ approximately.

Proof: According to Eq. (8), $\sum_{l=1}^{K-1}\frac{2l+1}{2l-1}\approx K-1$. Thus

$$T_M^L(K)\approx\frac{1}{1-p_{link}}(K-1)+\frac{1}{1-p_{link}}\lambda\pi(\frac{R}{K})^2 \quad (9)$$

By differentiating Eq. (9) and equating the result to 0 for minimizing $T_M^L(K)$ we can derive *Theorem 2.* ■

Next, the latency of SHCS and the minimal latency of KHCS ($K=K_{opt}^L$) is compared.

Proposition 4: Considering the latency of SHCS and KHCS (K>1), the latency of SHCS is longer than the minimal latency of KHCS ($K=K_{opt}^L$).

Proof: In this case, $T_M^L(K_{opt}^L)$ is changed to

$$T_M^L(K_{opt}^L)\approx\frac{1}{1-p_{link}}[(2\lambda\pi R^2)^{1/3}-1+\lambda\pi R^2\frac{1}{(2\lambda\pi R^2)^{2/3}}] \quad (10)$$

Compare Eq. (7) and Eq. (10), we have $T_S^L>T_M^L(K_{opt}^L)$. Thus the *Proposition 5* can be derived. ■

It is noted that the value of K_{opt}^L is derived from the viewpoint of latency in this section; comparably, the value of K_{opt}^E (given in Eq. (5)) is derived from the viewpoint of energy consumption in previous section. From Eq. (5) and *Theorem 2* we make out K_{opt}^E depends on the area of a cluster (R), while K_{opt}^L depends on the number of nodes in a cluster (λ and R).

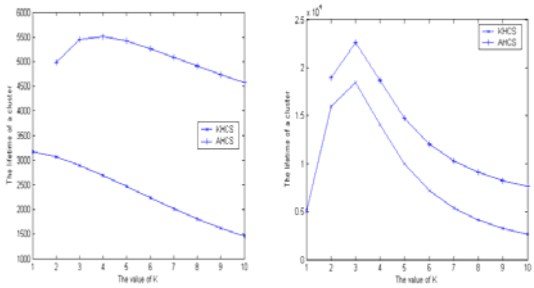
5. Simulations and Results

The simulations are executed to verify the correctness of analysis in previous Sections. All the results in this section are based on five runs with different random distribution of sensor nodes within a cluster. The radius of a cluster used for testing is 250m (R) and the value of p_{link} is 0.5 if not specified otherwise. The initial power for each sensor node is 2J. The data rate (r) for transmission is 50bit/second. Note for simplicity, we assume the connectivity of a cluster is assured in simulations.

5.1 Simulation I: Optimal Configuration for Energy consumption

When n=2 and n=4 the lifetime of a cluster is shown in Fig. 2(a) and Fig. 2(b), respectively.

From the Fig. 2(a), when n=2 we observe that the energy consumption of SHCS (K=1) is less than that of KHCS (K>1). From the Fig. 2(b), when n=4, the lifetime of KHCS is maximized if K(K_{opt}^E)=3, which proves the correctness of Eq. (5) (here R=250, α=50nJ/bit, β=0.001pJ/bit/m^4, n=4, p_{link}=0.5). Besides, Fig. 2(a) and (b) show the scheme AHCS excels KHCS and SHCS since it has longer lifetime within a cluster.



(a) (b)

Fig. 2 Value of K vs. the lifetime of a cluster: (a) n=2 (b) n=4.

5.2 Simulation II: Optimal Configuration for Latency

The latency of a cluster is shown in Fig. 3.

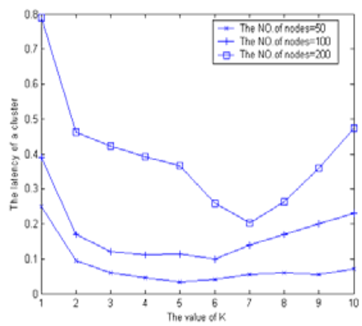


Fig. 3 Value of K vs. the latency of a cluster.

From Fig. 3, when the number of sensor nodes is 50, 100 and 200, the corresponding the optimal value of K (K_{opt}^{L}) is 5, 6 and 7 respectively. This numerical fact verifies the *Theorem 2*. Furthermore, SHCS scheme has the worst latency in all three cases. As a result, *Proposition 4* can be verified absolutely.

6. Conclusions

The optimal clustering scheme is very important for operators to execute periodical task efficiently. *Table 1~Table 3* summarize the optimal clustering scheme result in energy-centric and latency-centric application, respectively.

Table 1: Optimal clustering scheme in energy-centric application (n=2)

Parameter	SHCS	KHCS (K>1)
$R/\hat{r_c}=1$	√	X
$R/\hat{r_c}\geq 2$	√	K=2

Table 2: Optimal clustering scheme in energy-centric application (n=4)

Parameter		SHCS	KHCS (K>1)
$R/\hat{r_c}=1$		√	X
$R/\hat{r_c}\geq 2$	$K_{opt}^{E}=1<R/\hat{r_c}$	√	X
	$2\leq K_{opt}^{E}<R/\hat{r_c}$	√	$K=K_{opt}^{E}$
	$K_{opt}^{E}\geq R/\hat{r_c}$	√	$K=R/\hat{r_c}$

Table 3: Optimal clustering scheme in latency-centric application

Parameter		SHCS	KHCS (K>1)
$R/\hat{r_c}=1$		√	X
$R/\hat{r_c}\geq 2$	$K_{opt}^{L}=1<R/\hat{r_c}$	√	X
	$2\leq K_{opt}^{L}<R/\hat{r_c}$	X	$K=K_{opt}^{L}$
	$K_{opt}^{L}\geq R/\hat{r_c}$	X	$K=R/\hat{r_c}$

In *Table 1~3*, the character "X" denotes that the corresponding clustering scheme (here the scheme is KHCS) isn't to be considered. For example, assuming the application is energy-centric, if $R/\hat{r_c}\geq 2$ and $2\leq K_{opt}^{E}<R/\hat{r_c}$, then the scheme AHCS will be selected, i.e., the scheme SHCS and the scheme KHCS ($K=K_{opt}^{E}$) will rotate in turn periodically. In this case the lifetime of a cluster can be maximized. Therefore, under the configuring guideline in *Table 1~Table 3* the maximized lifetime or minimized latency can be achieved.

In summary, the key contributions of this paper are as follows:

- From the viewpoint of energy consumption, AHSC is the best scheme if KHCS (K>1) is feasible, and the value of K (K_{opt}^{E}) depends on the area of a cluster (R);
- From the viewpoint of latency, KHSC is the best scheme, and the value of K_{opt} (K_{opt}^{L}) is dependent on λ and R;
- The value of p_{link} leads to more energy required and longer latency for every node; it affects the optimal clustering scheme in the energy-centric application (Eq. (5)) instead of the latency-centric application (*Theorem 2*).
- The optimal clustering scheme (SHCS, KHCS (K>1) or AHCS) depends on the application requirement and other parameters (λ, R, p_{link} and $\hat{r_c}$).

There are other designing issues of sensor networks, such as finding the best-case path and worst-case path [14] by the approach of Voronoi diagram and Delaunay triangulation, as well as sleeping the number of nodes as many as possible to prolong the lifetime of sensor network, while not affect the quality of service, i.e., cover enough region and keeping the network connected [15].

In *Definition 1* of Section 2, the network lifetime is defined in terms of the time for the first node to die is often pessimistic, since it is very likely that the surviving nodes remain connected (or covered), thus not impairing the network functionality. In fact it is neither necessary nor desirable to have all nodes operate in the active mode at the same time. If all the sensor nodes simultaneously operated in the active mode, an excessive amount of energy would be wasted and data thus collected would be highly correlated and redundant. Moreover, excessive packet collision would occur as a result that many sensors intend to send packets especially in the presence of certain triggering events. As a result, how to design an optimal clustering scheme with minimal working nodes is our next step.

Another deficiency of this paper is that we assume the distribution of nodes is uniform, which maybe not always right. Supposing the sensor nodes will be spread from a plane. In this case, the density of the sensor nodes will depend on the velocity of plane. Certainly, the density of sensor nodes will increase if the velocity of plane is slow. Furthermore, designing an optimal clustering scheme with non-uniform sensor node deployment hasn't been discussed. Another problem is to investigate the influence of data aggregation on optimal clustering scheme. At last, although the unreliable wireless link is taken into account, the consideration about influence of failed senor node on the optimal clustering scheme is blank. These are left to the future direction.

Acknowledgments

The authors would like to thank the anonymous reviewers for their constructive comments and suggestions that greatly helped us improve the final quality of this paper.

References

[1] I. F. Akyildiz, Su. Weilian, Y. Sankarasubramaniam, and E.Cayirci, "A Survey on Sensor Networks," *IEEE Communications Magazine*, Vol. 40, No. 8, Aug. 2002, pp. 102−114

[2] E. Deborah, G. Ramesh, J. Heidemann, and S. Kumar, "Next Century Challenges: Scalable Coordination in Sensor Networks," in Proceedings of the ACM/IEEE International Conference on MOBICOM, Aug. 1999, pp. 263−270

[3] A. Cerpa, J. Elson, D. Estrin, L. Girod, M. Hamilton, and J. Zhao, "Habitat Monitoring: Application Driver for Wireless Communications Technology," in Proceedings of the ACM International Conference on SIGCOMM, Aug. 2001, pp. 20−41

[4] S. Eugene, C. Seong-Hwan, I. Nathan, M. Rex, S. Amit, A. Wang, and A. Chandrakasan, "Physical Layer Driven Protocol and Algorithm Design for Energy-Efficient Wireless Sensor Networks," in Proceedings of the ACM/IEEE International Conference on MOBICOM, Aug. 2001, pp. 272−287

[5] K.Arisha, M.Youssef, and M.Younis, "Energy-Aware TDMA-Based MAC for Sensor Networks," *Journal of Computer Networks (Elsevier)*, vol. 43, no. 5, Dec. 2003, pp. 539−694

[6] W. Heinzelman, J. Kulik, and H. Balakrishnan, "Adaptive Protocols for Information Dissemination in Wireless Sensor Networks," in Proceedings of the ACM/IEEE International Conference on MOBICOM, Aug. 1999, pp. 174−185

[7] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks," in Proceedings of the ACM/IEEE International Conference on MOBICOM, Aug. 2000, pp. 56−67

[8] V.Rodoplu and Teresa H. Meng, "Minimum Energy Mobile Wireless Networks," *IEEE Journal. Selected Areas in Communications*, Vol. 17, No. 8, Aug. 1999, pp. 1333−1344

[9] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Routing Protocols for Wireless Microsensor Networks," in Proceedings of 33rd Hawaii International Conference on System Sciences (HICSS '00), Jan. 2000, pp. 3005−3014

[10] K. Akkaya and M. Younis, "A Survey on Routing Protocols for Wireless Sensor Networks," *Journal of Ad hoc Networks (Elsevier)*, Vol. 3, No. 3, May. 2005, pp. 325−349

[11] B. Seema and J. C. Edward, "Minimizing Communication Costs in Hierarchically-clustered Networks of Wireless Sensors," *Journal of Computer Networks (Elsevier)*, Vol. 44, No. 1, Jan. 2004, pp. 1−16

[12] E. I. Oyman and E. Cem, "Overhead Energy Consideration for efficient routing in wireless sensor network," *Journal of Computer Networks*, Vol. 46, No. 4, Nov. 2004, pp. 465−478

[13] J. Pan, L. Cai, Y. Hou, Y. Shi, and X. Shen, "Optimal Base-Station Locations in Two-Tiered Wireless Sensor Networks," *IEEE Transactions on Mobile Computing*, vol. 4, no. 5, pp. 458−473, Sep. 2005

[14] X. Li, P. Wan, and O. Frieder, "Coverage in Wireless Ad Hoc Sensor Networks," *IEEE Trans. Computers*, Vol. 52, No. 6, June. 2003, pp. 753−763

[15] Z. Yi and K. Chakrabarty, "A Distributed Coverage and Connectivity-Centric Technique for Selecting Active Nodes in Wireless Sensor Networks," *IEEE Trans. Computers*, Vol. 54, No. 8, Aug. 2005, pp. 978−991

Yan Jin received the B.S. degree in Computer Science and Engineering from Harbin Institute of Technology, China in 2001, and the M.S. degree in Computer Science from Harbin Institute of Technology, China in 2003. He is currently pursuing the Ph.D. degree. His research interests include low-power architecture, topology management, performance evaluation and optimization designing in sensor networks, network simulations.