

Research Article

Distributed Joint Cluster Formation and Resource Allocation Scheme for Cooperative Data Collection in Virtual MIMO-Based M2M Networks

Xi Luan, Jianjun Wu, Yuxin Cheng, and Haige Xiang

Institution of Advanced Communications, EECS, Peking University, Beijing 100871, China

Correspondence should be addressed to Jianjun Wu; just@pku.edu.cn

Received 25 June 2014; Accepted 1 September 2014

Academic Editor: Yan Zhang

Copyright © 2015 Xi Luan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An efficient data collection scheme plays an important role for the real-time intelligent monitoring in many machine-to-machine (M2M) networks. In this paper, a distributed joint cluster formation and resource allocation scheme for data collection in cluster-based M2M networks is proposed. Specifically, in order to utilize the advantages of cooperation, we first propose a hierarchical transmission model which contains two communication phases. In the first phase, the intracluster information sharing is carried out by all the nodes within the same cluster. Then these nodes transmit the total information to the BS cooperatively with virtual-MIMO (VMIMO) protocol in the second phase. To grasp the properties and advantages of this cooperative transmission strategy, the theoretical analysis results are provided. The key issue in this system is to form the clusters and allocate resources efficiently. Since the optimization problem on this issue is an NP-hard problem, a feasible joint scheme for the cluster formation and resource allocation is proposed in this paper, which is carried out via coalition formation game with a distributed algorithm. This scheme can reduce the complexity while keeping an attractive performance. Simulation results show the properties of the proposed scheme and its advantages when comparing with the noncooperative scheme for the data collection in a practical scenario.

1. Introduction

Machine-to-machine (M2M) or machine type communications (MTC) have been recognized as a key and promising technique in next-generation communication networks [1]. Impelled by the recent theoretical advances, various kinds of M2M-based application networks are attracting more and more attentions, such as smart grid [2], vehicle communication networks [3], and e-health [4]. In these networks, various types of electronic terminals can be self-organized together and work without the intervention of humans. However, no matter in which applications, an efficient scheme for the real-time information collection plays an important role for the feasibility and reliability of the network, since it is crucial for the functions of network monitoring, automatic control, safety precautions, and so forth.

In many kinds of networks, the number of terminals may be very large, and their payload and distribution conditions are always different. Therefore, how to collect the real-time information from the whole network efficiently

and automatically is a great challenge. To deal with this issue, many traditional schemes can be utilized, such as the CSMA-based schemes, typical cellular multiple access schemes (e.g., CDMA, OFDMA) [5], or schemes with non-cooperative games [6]. However, with the limitation of frequency resources, as well as the growth of terminal number and payload size, cooperative schemes have showed more promising performance for the high efficient transmission with finite resources [7].

Among cooperative schemes, cluster-based communications are a promising research direction. Though it is attracting more and more attention, how to construct a highly efficient network structure and take full use of the limited resources is still a great challenge in M2M networks. The authors in [8] gave a comprehensive survey of the existing clustering algorithms for mobile ad hoc networks (MANETs). Literatures such as [9, 10] focused on the transmission schemes in cluster-based wireless sensor networks (WSN) under different situations. Furthermore, how to select or deploy the cluster heads has also been considered in the

context of cluster-based WSN networks [11]. However, the particularity of WSN makes these results not suitable for common M2M networks for generally two reasons. First, existing researches pay more attention on the battery life of sensors and take no account of the delay performance [10]. Second, the own properties of the information are always taken as the clustering criterion in WSN, such as compressibility [11]. Though some cooperative data collection schemes for M2M networks have also been studied recently, such as [12, 13], most of them focused on developing centralized algorithms. In addition, the joint cluster formation and resource allocation scheme, as well as the corresponding theoretical analysis, have not been sufficiently studied neither.

In this paper, we focus on cluster-based M2M network with VMIMO protocols. In this kind of networks, the members within a cluster transmit the information by means of VMIMO, which is different from the cooperative transmission schemes with cluster heads [14]. Though the VMIMO-based schemes have already been studied in WSN [15, 16], there were few practical solutions on the efficient cluster formation by taking resource allocation into consideration. Meanwhile, the quantitative theoretical analysis on the advantages and properties of this kind of system has also not been given clearly in existing researches.

In this context, a distributed joint cluster formation and resource allocation scheme is proposed in this paper for the real-time data collection in VMIMO-based M2M networks. We first give a hierarchical cooperative transmission scheme with two communication phases: the information sharing phase and the cooperative transmission phase. To demonstrate the advantages of this cooperative scheme and the influence of different factors on performance, some theoretical results are given. Based on these work, we propose a joint cluster formation and resource allocation scheme based on coalition formation game and then provide the whole implementation procedure via a distributed cooperative algorithm. In this paper, time cost for data collection is chosen to be the performance indicator to meet the requirement of some real-time applications.

The rest of this paper is organized as follows. We first present the system model in Section 2. Then an optimization problem and the corresponding theoretical analysis are given in Section 3. In Section 4, a distributed joint cluster formation and resource allocation scheme is proposed. Simulation results and the data analysis are provided in Section 5. Finally, we conclude this paper in Section 6.

2. System Model

Consider a cooperative data collection system in M2M networks shown in Figure 1. There are totally M distributed single-antenna data acquisition nodes (DANs), a base station (BS) with N_R antennas, and a data center within a certain region. The roles of these nodes are different in different networks, such as smart meters in smart grid or vehicle sensors in vehicle communications networks. All the information acquired by these DANs should be collected by the BS timely and forwarded to the data center (such as power center in smart grid).

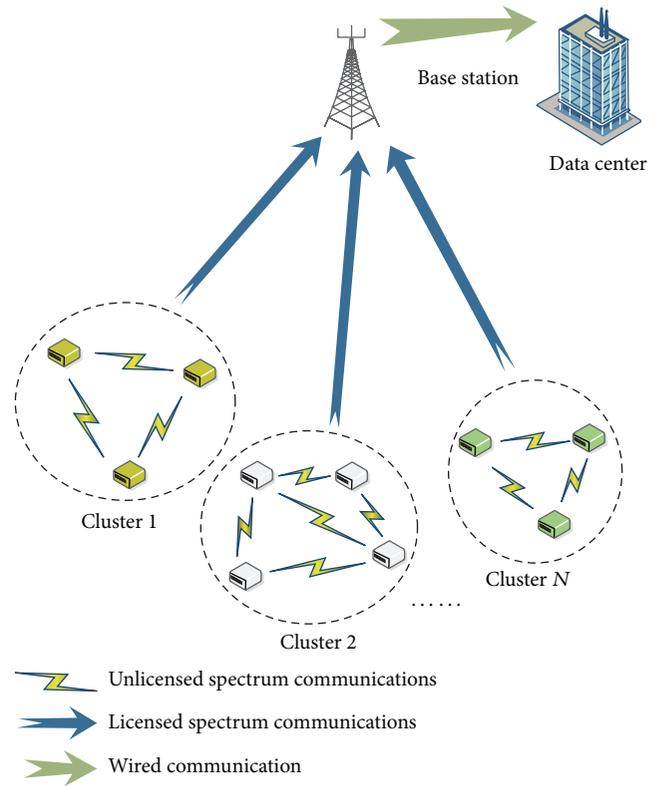


FIGURE 1: An illustration of system model.

2.1. Cluster-Based Data Collection. In this system, the data collection process can be carried out cooperatively and hierarchically. Before transmission, the DANs can organize themselves into N ($1 \leq N \leq M$) disjoint clusters according to certain rules without the intervention of humans. Let $\mathcal{S} = \{\mathcal{S}_1, \dots, \mathcal{S}_N\}$ be the set of clusters, where \mathcal{S}_j is the j th cluster with $|\mathcal{S}_j|$ nodes. Based on the cluster-organized network structure, the whole data collection process can be divided into two phases (subprocesses). In the first phase, the nodes in the same cluster share their data information $\{x_{1,j}, \dots, x_{|\mathcal{S}_j|,j}\}$ with each other by means of broadcasting, where $x_{i,j}$ denotes the payload of the i th DAN in cluster \mathcal{S}_j . Since the range of these intracluster communications is always short, unlicensed spectrum can be utilized to enhance the spectral efficiency. In the second phase, the nodes in the same cluster cooperate together as a virtual MIMO (VMIMO) transmitter and send the combined payload $x_{\mathcal{S}_j} = \sum_{i \in \mathcal{S}_j} x_{i,j}$ to the BS jointly. In this phase, licensed spectrum (such as the cellular spectrum) is occupied for the possibility of long-distance communication. In the first-layer transmission, TDMA scheme is chosen, since it is the most common option for the nodes to support various kinds of short-distance communication approaches in different scenarios, such as WiFi, Zigbee, and bluetooth. However, since the distances in the second-layer transmission are much larger, it is necessary for them to adopt a compatible MA scheme with LTE system to avoid extra interference. Therefore, we adopt OFDMA scheme for the second-layer transmission.

2.2. Traffic Model. The payload conditions are different in different kinds of networks. In this paper, we define the traffic model according to IEEE Std 2030 [17], which is the guidance provided by IEEE to model the features of smart grid. For the i th DAN, the packet number and the size of each packet all obey truncated Pareto distribution, which has also been used in several data analysis fields [17]. Assume the upper and lower bounds of packet number are n_{\max} and n_{\min} ; the packet number of the i th DAN can be generated with the following probability density function correspondingly:

$$f(n_i) = \begin{cases} \frac{\kappa_1 \cdot n_{\min}^{\kappa_1} \cdot (n_i)^{-\kappa_1-1}}{1 - (n_{\min}/n_{\max})^{\kappa_1}}, & n_{\min} < n_i \leq n_{\max}, \\ 0, & \text{others,} \end{cases} \quad (1)$$

where κ_1 is the constant parameter of Pareto distribution and $(1 - (n_{\min}/n_{\max})^{\kappa_1})^{-1}$ is the probability normalization factor. Similarly, the length of each packet ranges from s_{\max} to s_{\min} also obeys the truncated Pareto distribution with parameter κ_2 .

Hence, the total payload of the i th DAN is $x_i = \sum_{\xi=1}^{n_i} s_i^\xi$, which is to be shared in the first communication phase within cluster \mathcal{S}_j .

2.3. Data Transmission Model. As mentioned above, the data collection process in the system contains two phases, and hence the data transmission model also can be formed with two parts.

Since the TDMA scheme is adopted in the intracluster information sharing process, all DANs within the same cluster broadcast their information sequentially. We assume that the transmit power of DANs can be low enough to eliminate the intercluster interference (ICI). In this context, the minimum time cost for the i th DAN to share its information with all the others in the same cluster \mathcal{S}_j can be expressed as

$$T_{i,j} = \frac{x_{i,j}}{W \cdot \log\left(1 + P_1 |h_{i,\hat{i}}|^2 d_{i,\hat{i}}^{-\beta} / \sigma^2\right)}, \quad (2)$$

where $h_{i,\hat{i}}$ and $d_{i,\hat{i}}$ are the channel and distance between the i th DAN and its farthest partner \hat{i} in cluster \mathcal{S}_j , β is the path loss factor, W is the available bandwidth in unlicensed spectrum, P_1 is the transmit power of DANs in the first communication phase, and σ^2 is the noise variance.

Therefore, the total time cost for finishing the information sharing process in cluster \mathcal{S}_j can be calculated as

$$T_{\mathcal{S}_j}^{(1)} = \sum_{i \in \mathcal{S}_j} T_{i,j}. \quad (3)$$

Before cooperative transmission, the DANs in the same cluster combine their own data with the received data and generate a whole payload $x_{\mathcal{S}_j} = \sum_{i \in \mathcal{S}_j} x_{i,j}$. After that, they transmit $x_{\mathcal{S}_j}$ to the BS with VMIMO protocol in the second phase. In this phase, OFDMA scheme is adopted. Hence, after

being multiplied by weight vector, the received signal at the BS that comes from cluster \mathcal{S}_j in the second phase is

$$\bar{y}_{\mathcal{S}_j} = \sqrt{\frac{P_2 M_j}{d_{\mathcal{S}_j}^\beta}} \mathbf{w}_{R,\mathcal{S}_j}^H \mathbf{H}_{\mathcal{S}_j} \mathbf{w}_{T,\mathcal{S}_j} x_{\mathcal{S}_j} + \bar{n}_{\mathcal{S}_j}, \quad (4)$$

where $M_j = |\mathcal{S}_j|$ is the number of DANs in \mathcal{S}_j , $\mathbf{H}_{\mathcal{S}_j} \in \mathbb{C}^{N_R \times M_j}$ is the channel matrix among DANs and antennas of the BS, $\mathbf{w}_{T,\mathcal{S}_j} \in \mathbb{C}^{M_j \times 1}$ is the normalized precoding vector adopted by DANs in \mathcal{S}_j , that is, $E[\mathbf{w}_{T,\mathcal{S}_j}^H \mathbf{w}_{T,\mathcal{S}_j}] = \mathbf{I}_{M_j}$, P_2 is the transmit power in the second phase, $d_{\mathcal{S}_j}$ is the distance between the cluster center and the BS, $\mathbf{w}_{R,\mathcal{S}_j} \in \mathbb{C}^{N_R \times 1}$ is the receiving weight vector of the BS, $n_{\mathcal{S}_j} = \mathbf{w}_{R,\mathcal{S}_j}^H \mathbf{n}_{\mathcal{S}_j}$ is the scalar noise after processing, and $\mathbf{n}_{\mathcal{S}_j}$ is the additive white Gaussian noise (AWGN), denoted by $\mathcal{CN}(0, \sigma^2)$.

In this system, maximal ratio transmission (MRT) and maximal ratio combining (MRC) are adopted by DANs and the BS, respectively [18]. Specifically, the precoding vector for DANs in \mathcal{S}_j is $\mathbf{w}_{T,\mathcal{S}_j} = \mathbf{v}_{1,\mathcal{S}_j}$, where $\mathbf{v}_{1,\mathcal{S}_j}$ is the first right singular vector of the channel matrix $\mathbf{H}_{\mathcal{S}_j}$. Correspondingly, the receiving weight vector for the BS is $\mathbf{w}_{R,\mathcal{S}_j} = \mathbf{H}_{\mathcal{S}_j} \cdot \mathbf{v}_{1,\mathcal{S}_j}$. In this case, the signal-to-noise ratio (SNR) is

$$\gamma_{\mathcal{S}_j} = \frac{P_2 M_j \mu_{\max,\mathcal{S}_j}^2}{d_{\mathcal{S}_j}^\beta \sigma^2}, \quad (5)$$

where μ_{\max,\mathcal{S}_j} is the largest singular value of $\mathbf{H}_{\mathcal{S}_j}$, which corresponds to the first right singular vector $\mathbf{v}_{1,\mathcal{S}_j}$.

Then the time cost for cooperative transmission in the second phase can be derived as

$$T_{\mathcal{S}_j}^{(2)} = \frac{x_{\mathcal{S}_j}}{w_{\mathcal{S}_j} \cdot \log\left(1 + \gamma_{\mathcal{S}_j}\right)}, \quad (6)$$

where $w_{\mathcal{S}_j}$ is the available bandwidth of cluster \mathcal{S}_j .

Therefore, the total time cost for the BS to collect all data from cluster \mathcal{S}_j is

$$T_{\mathcal{S}_j} = T_{\mathcal{S}_j}^{(1)} + T_{\mathcal{S}_j}^{(2)}. \quad (7)$$

Obviously, the total time cost for information submission of the i th DAN in cluster \mathcal{S}_j is also equal to $T_{\mathcal{S}_j}$; that is, $T_{i,\mathcal{S}_j} = T_{\mathcal{S}_j}$.

Since in many application scenarios, time-effectiveness is a key performance indicator or even determines the success or failure of network functions, such as network monitoring or safety precaution in vehicle networks or smart grid, we will take time cost as the performance evaluation criterion in this paper. As we can see from (2) to (7), the network structure and resource allocation strategy in this system can influence the performance. In next section, the theoretical analysis will be given to show this influence explicitly. After that, a practical scheme will be proposed.

3. Theoretical Analysis on Hierarchical Cooperative Transmission

3.1. Optimization Problem Formulation. In this system, the ideal solution is to find out a high-efficient network structure and a resource allocation strategy to minimize the average time cost for data collection, which is defined as follows:

$$\bar{T} = \frac{1}{M} \sum_{j=1}^N \sum_{i=1}^{M_j} T_{i,\mathcal{S}_j} = \frac{1}{M} (M_1 T_{\mathcal{S}_1} + \cdots + M_N T_{\mathcal{S}_N}). \quad (8)$$

If the global network information is available at the BS, such as the channel information, the payload condition, and the distribution of DANs, a centralized approach can be used to obtain the optimal system performance.

To describe the cooperative relationships, a network structure indicator matrix $\mathbf{S} \in \mathbb{C}^{M \times N}$ and a resource allocation indicator matrix $\mathbf{W} \in \mathbb{N}^{N \times K}$ can be defined, where N and K are the cluster number and available subband number, respectively. Their elements are

$$\begin{aligned} s_{i,j} &= 1, & \text{if } i \in \mathcal{S}_j, \\ s_{i,j} &= 0, & \text{if } i \notin \mathcal{S}_j, \\ w_{j,k} &= 1, & \text{if } k \in \mathbf{w}_j, \\ w_{j,k} &= 0, & \text{if } k \notin \mathbf{w}_j, \end{aligned} \quad (9)$$

where \mathbf{w}_j indicated the available subbands for cluster \mathcal{S}_j . Thus, the time cost $T_{\mathcal{S}_j}$ can be rewritten as

$$T_{\mathcal{S}_j} = \sum_{i=1}^M s_{i,j} T_{i,j} + \frac{\sum_{i=1}^M s_{i,j} x_{i,j}}{\sum_{k=1}^K w_{j,k} \cdot \log(1 + \gamma_{j,k})}. \quad (10)$$

The average time cost \bar{T} in (8) can also be redefined correspondingly. In this context, the following optimization problem can be formulated:

$$\begin{aligned} & \underset{N, \mathbf{S}, \mathbf{W}}{\text{Minimize}} && \bar{T}, \\ & \text{s.t.} && 1 \leq N \leq M, \\ & && T_{\mathcal{S}_j} \leq T_0, \\ & && \sum_{j=1}^N s_{i,j} = 1, \\ & && \sum_{j=1}^N w_{j,k} = 1, \end{aligned} \quad (11)$$

where T_0 is the constraint on the maximum period. The third constraint guarantees the nonoverlapping cluster formation, which means that one DAN can only join in one cluster. The last constraint ensures the exclusive bandwidth allocation to avoid interference among different clusters.

However, problem (11) is a nonlinear integer programming problem. To obtain the optimal solution, all possibilities of cluster formation and the corresponding optimal resource allocation pattern should be enumerated, which is an NP-hard problem when the DAN number grows up. Though the optimal solutions can hardly be obtained, the properties of them and the influence factors on system performance can be obtained by the theoretical analysis, which is our work in the next subsection.

3.2. Advantages of Hierarchical Cooperative Transmission. Consider the cluster \mathcal{S}_j with M_j DANs. To prove the advantages of cluster-based transmission with VMIMO protocol, we compare the average time cost of the noncooperative and cooperative schemes under the same condition of bandwidth and power resources. Assume that the available bandwidth is w_j and the transmit power of each node is P_2 ; then if all DANs send their information separately with time division strategy and the BS processes the received signals with MRC, the average time cost can be calculated as follows:

$$\begin{aligned} \bar{T}_{j,s} &= E \left(\frac{x_i}{w_j/M_j \cdot \log(1 + P\mu_{i,\max}^2 d_i^{-\beta}/\sigma^2)} \right) \\ &= \frac{1}{M_j} \sum_{i=1}^{M_j} \frac{E(x_i)}{w_j/M_j \cdot \log(1 + P\mu_{i,\max}^2 d_i^{-\beta}/\sigma^2)} \\ &= \frac{E(n)E(s)}{w_j} \left(\sum_{i=1}^{M_j} \frac{1}{\log(1 + P\mu_{i,\max}^2 d_i^{-\beta}/\sigma^2)} \right), \end{aligned} \quad (12)$$

where $\mu_{i,\max}$ is the largest singular value of the channel vector $\mathbf{h}_i \in \mathbb{C}^{N_R \times 1}$, d_i is the distance between node i and the BS, and $E(n)$ and $E(s)$ are the average packet number and packet size of DANs, respectively. Since both n and s obey truncated Pareto distribution, $E(n)$ and $E(s)$ can be calculated as follows:

$$\begin{aligned} E(n) &= \frac{n_{\min}^{\kappa_1}}{1 - (n_{\min}/n_{\max})^{\kappa_1}} \cdot \frac{\kappa_1}{\kappa_1 - 1} \cdot \frac{n_{\max}^{\kappa_1-1} - n_{\min}^{\kappa_1-1}}{n_{\min}^{\kappa_1-1} \cdot n_{\max}^{\kappa_1-1}}, \\ E(s) &= \frac{s_{\min}^{\kappa_2}}{1 - (s_{\min}/s_{\max})^{\kappa_2}} \cdot \frac{\kappa_2}{\kappa_2 - 1} \cdot \frac{s_{\max}^{\kappa_2-1} - s_{\min}^{\kappa_2-1}}{s_{\min}^{\kappa_2-1} \cdot s_{\max}^{\kappa_2-1}}. \end{aligned} \quad (13)$$

According to (6), if the DANs in cluster \mathcal{S}_j cooperate with each other as a VMIMO system, the average time cost can be written as

$$\bar{T}_j^* = \frac{M_j E(n) E(s)}{w_j} \left(\frac{1}{\log(1 + M_j P_2 \mu_{\max}^2 d^{-\beta}/\sigma^2)} \right), \quad (14)$$

where μ_{\max} is the largest singular value of the whole channel matrix $\mathbf{H}_j = [\mathbf{h}_1 \cdots \mathbf{h}_{M_j}]$, d is the average distance between DANs and the BS, and $E(n)$ and $E(s)$ are the same with the ones defined in (13).

Since \mathbf{H}_j is composed of $\mathbf{h}_i, \forall i \in \mathcal{S}_j$, the relationship $\mu_{\max} > \mu_{i,\max}, \forall i \in \mathcal{S}_j$ is always true according to matrix theory. Assume that the range of a cluster is much smaller than the distance to the BS; then the approximation $d \approx \bar{d}_i$ holds. Define $R_i \triangleq \log(1 + P_2 \mu_{i,\max}^2 \bar{d}_i^{-\beta} / \sigma^2)$ and $R^* \triangleq \log(1 + M_j P_2 \mu_{\max}^2 \bar{d}^{-\beta} / \sigma^2)$ as the achievable rates in unit bandwidth; then difference value between $\bar{T}_{j,s}$ and \bar{T}_j^* can be calculated as

$$\begin{aligned} \bar{T}_{j,s} - \bar{T}_j^* &= \frac{E(n)E(s)}{w_j} \left(\sum_{i=1}^{M_j} \frac{1}{R_i} - \frac{M_j}{R^*} \right) \\ &= \frac{E(n)E(s)}{w_j} \frac{R^* \left(\sum_{i=1}^{M_j} \prod_{k=1, k \neq i}^{M_j} R_k \right) - M_j \prod_{i=1}^{M_j} R_i}{\prod_{i=1}^{M_j} R_i \cdot R^*} \\ &= \frac{E(n)E(s)}{w_j} \frac{\sum_{i=1}^{M_j} (R^* - R_i) \prod_{k=1, k \neq i}^{M_j} R_k}{\prod_{i=1}^{M_j} R_i \cdot R^*}. \end{aligned} \quad (15)$$

Since $R^* > R_i, \forall i \in \mathcal{S}_j$, then $\bar{T}_{j,s} - \bar{T}_j^* > 0$ holds. From (15) it can be seen that, for a certain number of DANs, the proposed cluster-based transmission scheme can bring more benefit on transmission efficiency when comparing with noncooperative scheme under the same condition.

However, the result in (15) has not taken the cluster formation cost into consideration. Actually, for a whole data collection process as mentioned in Section 2, the cost in information sharing phase also cannot be ignored. According to (2) and (3), the average time cost for the information sharing process in cluster \mathcal{S}_j is

$$\bar{T}_j^{(1)} = \frac{E(n)E(s)}{W} \left(\sum_{i=1}^{M_j} \frac{1}{\log(1 + P_1 |h_{i,\hat{i}}|^2 \bar{d}_{i,\hat{i}}^{-\beta} / \sigma^2)} \right). \quad (16)$$

It shows that this process not only increases the time cost for data collection but also occupies a certain amount of power resources and therefore brings negative effect to the system performance. Define $\tilde{R}_i = \log(1 + P_1 |h_{i,\hat{i}}|^2 \bar{d}_{i,\hat{i}}^{-\beta} / \sigma^2)$. Combining (15) and (16), we can calculate the final performance enhancement for this scheme, which is given as follows:

$$\begin{aligned} \Delta \bar{T}_j &= E(n)E(s) \\ &\times \left\{ \left(W \cdot \prod_{i=1}^{M_j} \tilde{R}_i \cdot \left[\sum_{i=1}^{M_j} (R^* - R_i) \prod_{k=1, k \neq i}^{M_j} R_k \right] \right. \right. \\ &\quad \left. \left. - w_j \cdot \prod_{i=1}^{M_j} R_i \cdot R^* \cdot \left[\sum_{i=1}^{M_j} \prod_{k=1, k \neq i}^{M_j} \tilde{R}_k \right] \right) \right. \\ &\quad \left. \times \left(w_j \cdot W \cdot \prod_{i=1}^{M_j} (\tilde{R}_i R_i) \cdot R^* \right)^{-1} \right\}. \end{aligned} \quad (17)$$

The result in (17) shows that the performance enhancement is influenced by many factors. And in all of these factors, the network structure and the resource allocation strategy are the most important ones.

3.3. Influence of Network Structure and Resource Allocation. As shown in (17), there are many parameters related to the network structure, such as $M_j, h_{i,\hat{i}}, \mathbf{H}_j$, and $\bar{d}_{i,\hat{i}}$. Therefore, how to form an optimal network structure is a key issue in this network. Define the network structure as $\mathcal{S} = \{\mathcal{S}_1, \dots, \mathcal{S}_N\}$, and the following property can be established.

Property 1. Though the cluster-based cooperative transmission scheme can bring benefit for the system, a grand cluster, which is denoted by $\mathcal{S} = \mathcal{M}$, can hardly be formed to get the optimal performance.

Proof. Assume that the total bandwidth for the second phase transmission is W_2 , and the available bandwidth of cluster \mathcal{S}_j is $w_j = M_j \cdot W_2 / M$. Consider the average time cost in the cooperative transmission phase \bar{T}_j^* , and if the DAN number in cluster \mathcal{S}_j increases, the first-order and second-order derivatives can be calculated as

$$\begin{aligned} \frac{\partial \bar{T}_j^*}{\partial M_j} &= -\frac{ME(n)E(s)}{W_2(R^*)^2} \frac{\eta}{1 + \eta M_j} < 0, \\ \frac{\partial^2 \bar{T}_j^*}{\partial^2 M_j} &= \frac{ME(n)E(s)}{W_2(R^*)^2} \frac{\eta^2}{(1 + \eta M_j)^2} > 0, \end{aligned} \quad (18)$$

where $\eta \triangleq P_2 \mu_{\max}^2 \bar{d}^{-\beta} / \sigma^2$. From (18), it can be seen that, with the increase of M_j , the \bar{T}_j^* can be reduced. However, the declining trend of \bar{T}_j^* becomes more and more slow. On the other hand, the average farthest distance in a cluster, $\bar{d}_{i,\hat{i}} = (1/M_j) \sum_{i \in \mathcal{S}_j} \bar{d}_{i,\hat{i}}$, is generally proportional to M_j ; that is, $\bar{d}_{i,\hat{i}} \propto M_j$. Meanwhile, in order to avoid the interference among clusters, the transmit power P_1 should be low enough. Therefore, the first-order derivative $\partial \bar{T}_j^{(1)} / \partial M_j > 0$, which means that the average time cost for the information sharing in the first phase scales up with M_j .

Thus, the average total time cost $\bar{T}_j = \bar{T}_j^{(1)} + \bar{T}_j^*$ is a convex function of M_j . With the growth of M_j , \bar{T}_j may even become larger than the time cost $\bar{T}_{j,s}$ of noncooperative scheme. Then, some DANs have incentive to deviate from the grand cluster and form small clusters instead. Therefore, the grand cluster can hardly be formed. \square

Apart from network structure, power and frequency resource allocation can also impact the performance greatly. Assume that the total available power for a DAN is P_0 , which can be allocated to the two communication phases; that is, $P_0 = P_1 + P_2$, as shown in (2) and (5). Though it can be verified that for the total time cost in (7), $\partial^2 T_{\mathcal{S}_j} / \partial^2 P_1 > 0$, the optimal point of power allocation cannot be chosen because interference among clusters may be introduced if P_1 is large.

Therefore, for simplicity, we set $P_1 = \alpha P_2$ in this system, where α is a constant satisfying $0 < \alpha < 0.5$.

Since the time-division scheme is adopted for the intra-cluster transmission, the frequency resource allocation refers to the bandwidth allocation in the second communications phase. According to (14) and (17), it can be calculated that

$$\begin{aligned} \frac{\partial \bar{T}_j^*}{\partial w_j} &= -\frac{M_j E(n) E(s)}{w_j^2 R^*} < 0, \\ \frac{\partial \Delta \bar{T}_j}{\partial w_j} &= -\frac{E(n) E(s)}{w_j^2} \frac{\sum_{i=1}^{M_j} (R^* - R_i) \prod_{k=1, k \neq i}^{M_j} R_k}{\prod_{i=1}^{M_j} R_i \cdot R^*} < 0. \end{aligned} \quad (19)$$

The results in (19) show that, for a certain cluster, with the increase of available bandwidth, the average total time cost $\bar{T}_j = \bar{T}_j^{(1)} + \bar{T}_j^*$ can be reduced. On the other hand, the average performance enhancement $\Delta \bar{T}_j$ also decreases with the growth of available bandwidth, which means the decrease of spectral efficiency. Therefore, an appropriate resource allocation strategy is necessary for the network to maximize the spectral efficiency and meanwhile guarantee the fairness among clusters.

According to the above theoretical analysis, a joint cluster formation and resource allocation scheme will be proposed in the next section, which is implemented via distributed method. Though the time cost performance of the proposed scheme may be worse than the centralized scheme, the optimal solution of the later scheme can hardly be obtained in practical M2M applications, since it is hard to deploy a centric controller for all the DANs and meanwhile the computing complexity is quite high.

4. Distributed Joint Cluster Formation and Resource Allocation Scheme

In this section, we first formulate the problem as a coalitional game. Then a simplified resource allocation criterion is provided to transfer the complicated original game into a feasible game. After that, we give the detailed implementation process for the scheme based on a practical distributed algorithm.

4.1. Coalition Formation Game Formulation. We formulate the cooperation among DANs as a coalitional game $(\mathcal{M}, \mathcal{V})$, where \mathcal{M} is the player set who are seeking partners to maximize their utilities, and \mathcal{V} is the payoff mapping of these players. For a coalition \mathcal{S}_j , if its utility function $v(\mathcal{S}_j)$ is only decided by its own members and irrelevant with the player set $\mathcal{M} \setminus \mathcal{S}_j$, it is a coalitional game with characteristic form [19]. On the contrary, if the behaviors of players outside can influence $v(\mathcal{S}_j)$, the game can be recognized as a coalitional game with partition form. On the other hand, from the view of utility allocation, coalitional game can be divided into two categories: games with transferable utility (TU) and nontransferable utility (NTU). In the former category, the coalition utility $v(\mathcal{S}_j)$ can be divided and transferred among

players inside, which satisfies $\sum_{i \in \mathcal{S}_j} v_i = v(\mathcal{S}_j)$, where v_i is the payoff of the i th player. In NTU coalitional game, however, the utility $v(\mathcal{S}_j)$ is a payoff vector, whose elements are the payoffs of players decided by their own strategy and joint behaviors of others in \mathcal{S}_j .

In this system, the utility function of the i th DAN in cluster \mathcal{S}_j is defined as

$$v_{i,j} = -T_{i,\mathcal{S}_j} = -\left(T_{\mathcal{S}_j}^{(1)} + T_{\mathcal{S}_j}^{(2)}\right), \quad (20)$$

where T_{i,\mathcal{S}_j} has been defined in (7). Since the time costs of all DANs in the same cluster are the same, the total utility of cluster \mathcal{S}_j is

$$v(\mathcal{S}_j) = \sum_{i \in \mathcal{S}_j} v_{i,j} = -M_j \cdot \left(T_{\mathcal{S}_j}^{(1)} + T_{\mathcal{S}_j}^{(2)}\right). \quad (21)$$

Among all coalitional games, canonical coalitional games and coalition formation games are the most typical ones which have been studied in communication systems [20]. However, since we have already proved that the grand cluster cannot be formed in Section 3, we only focus on the later kind of games, in which many small clusters can be formed to benefit all DANs in this network.

As we can see, the utility function $v(\mathcal{S}_j)$ contains the term w_j , which is not only determined by DANs in cluster \mathcal{S}_j but also influenced by the resource allocation strategy of the whole network. Therefore, this game can be recognized as a NTU coalition formation game (\mathcal{M}, v) with partition mode. However, coalitional games with partition mode are always quite complicated to be solved in practical scenario [19]. In the next subsection, we will provide a simplified resource allocation rule in this system with the aim to transfer the game into the one with characteristic mode.

4.2. Resource Allocation. In this system, both power and frequency resources can be allocated. As shown in Section 3, though there is an optimal tradeoff between $T_{\mathcal{S}_j}^{(1)}$ and $T_{\mathcal{S}_j}^{(2)}$ with proper power allocation, we set a small fixed ratio between P_1 and P_2 to avoid the intercluster interference for simplicity. Therefore, the resource allocation here refers to the bandwidth allocation in the second phase.

In order to transfer the original game into a game with characteristic mode, the available bandwidth $w_{\mathcal{S}_j}$ for cluster \mathcal{S}_j should only depend on its own members. To achieve this goal, we first divide the total bandwidth W_2 into M parts virtually with the same size $w_v = W_2/M$. Then, each DAN is assumed to be bound with one part of the bandwidth resource. If the i th DAN joins into a cluster \mathcal{S}_j , the bandwidth $w_{i,v}$ can be used by all DANs within the same cluster. In this way, the frequency resource can be recognized as being carried out in nodes' level and therefore the total available bandwidth for cluster \mathcal{S}_j is $w_j = M_j \cdot w_v$. Though it is not the optimal allocation, the computing complexity of this manner is much lower when comparing with the centralized scheme shown in (11). Meanwhile, $v(\mathcal{S}_j)$ here is only determined by the DANs in \mathcal{S}_j and therefore the original game has been transferred into a game with characteristic mode.

Initialization

Set each individual DAN as a cluster, and the initial partition is $\mathcal{S} = \mathcal{M} = \{\mathcal{S}_1, \dots, \mathcal{S}_M\}$.

Joint cluster formation and virtual resource allocation**while**

- (1) Merge the subset $\{\mathcal{S}_{j_1}, \dots, \mathcal{S}_{j_k}\}$ into \mathcal{S}_j if the merge rule satisfies. Meanwhile, re-allocate the combined virtual bandwidth resources $\{w_{j_1,v}, \dots, w_{j_k,v}\}$ to the new cluster, $w_{j,v} = \sum_{i \in \{j_1, \dots, j_k\}} w_{i,v}$.
- (2) Split \mathcal{S}_j into a set of small clusters $\{\mathcal{S}_{j_1}, \dots, \mathcal{S}_{j_k}\}$ if the split rule satisfies. Meanwhile, re-allocate the bandwidth resource $w_{j,v}$ to all split clusters proportionally according to the DAN number in each cluster, $w_{\mathcal{S}_k,v} = (M_k / \sum_{i \in \{j_1, \dots, j_k\}} M_i) w_{j,v}$.

until

A stopping criterion is satisfied.

Result reported and final resource allocation

- (1) Report the final partition $\mathcal{S}^* = \{\mathcal{S}_{1,v}^*, \dots, \mathcal{S}_{N,v}^*\}$ and the corresponding virtual resource allocation results $\mathcal{W} = \{w_{1,v}^*, \dots, w_{N,v}^*\}$ to the BS.
- (2) The BS allocates proper frequency resources for each cluster with the same amount denoted by \mathcal{W} .

ALGORITHM 1: Implementation of the joint cluster formation and resource allocation scheme.

4.3. Distributed Implementation for Joint Cluster Formation and Resource Allocation. According to the theoretical results in Section 3 and the resource allocation rule above, the joint problem of cluster formation and resource allocation can be solved with a distributed method. Before the description of the implementation process, some concepts should be briefly introduced [19].

Definition 1 (partition). If the whole set of DANs \mathcal{M} can be divided into a set of coalitions $\mathcal{S} = \{\mathcal{S}_1, \dots, \mathcal{S}_N\}$, which satisfies $\cup_{j=1}^N \mathcal{S}_j = \mathcal{M}$ and $\mathcal{S}_j \cap \mathcal{S}_k = \emptyset, \forall j \neq k$, \mathcal{S} can be defined as a partition of \mathcal{M} . Each partition represents a kind of network structure.

Definition 2 (utilitarian order). Assume that $\mathcal{R} = \{R_1, \dots, R_L\}$ and $\mathcal{S} = \{\mathcal{S}_1, \dots, \mathcal{S}_N\}$ are two different partitions of \mathcal{M} . The DANs prefer to self-organize into the clusters indicated by partition \mathcal{S} with utilitarian order if the following condition is satisfied:

$$\mathcal{S} \triangleright \mathcal{R} \iff \sum_{j=1}^N v(\mathcal{S}_j) > \sum_{j=1}^L v(\mathcal{R}_j). \quad (22)$$

The operator \triangleright shows the comparison relation. Equation (22) means that the total social welfare in the network increases if the network structure changes from \mathcal{R} to \mathcal{S} . In other words, the average time cost for data collection with network structure \mathcal{S} is lower.

Based on these definitions, two important rules for the partition adjustment can be introduced [19], which are the foundation of the distributed implementation process that will be described later.

- (i) **Merge Rule.** For any subset of clusters $\{\mathcal{S}_{j_1}, \dots, \mathcal{S}_{j_k}\}$, they can merge to a new cluster \mathcal{S}_j when the condition $\{\mathcal{S}_j\} = \{\cup_{i=j_1}^{j_k} \mathcal{S}_{j_i}\} \triangleright \{\mathcal{S}_{j_1}, \dots, \mathcal{S}_{j_k}\}$ is satisfied.
- (ii) **Split Rule.** For any cluster \mathcal{S}_j , it can be split into a set of small clusters $\{\mathcal{S}_{j_1}, \dots, \mathcal{S}_{j_k}\}$ if the condition $\{\mathcal{S}_{j_1}, \dots, \mathcal{S}_{j_k}\} \triangleright \{\mathcal{S}_j\} = \{\cup_{i=j_1}^{j_k} \mathcal{S}_{j_i}\}$ is satisfied.

This merge-and-split algorithm has also been used in other cooperative networks [20]. Based on this algorithm, the implementation process for the distributed joint cluster formation and resource allocation scheme can be designed. The detailed procedures are provided in Algorithm 1. The whole process can be carried out in a distributed method and therefore is easy to be achieved. Specifically, in the iteration process, all DANs try to find their proper partners to get the maximum utilities. The solution of network structure and resource allocation changes virtually with each merge and split operation. Only when the iteration process stops, the final solution of network structure and resource allocation is valid and can be executed actually by the DANs and the BS.

Authors in [19] have proved that the merge-and-split algorithm can converge to at least one kind of stability, \mathbb{D}_{hp} . Since the cluster utilities keep increasing in the process of merge-and-split operations, the final results can be regarded as optimal solutions in the view of utilitarian order. Meanwhile, since there is no payload transmission in the virtual iteration process, the time cost for cluster formation is very low. Therefore, though the performance of the proposed scheme is worse than the optimal solution in problem (11), the computing and implementing complexity is much lower while the solutions in (11) can hardly be obtained.

5. Simulation Results

To verify the performance of the proposed scheme, we set up a practical simulation scenario. A number of DANs are uniformly distributed within a 1 km \times 1 km square area and the BS is located at the center of this focused area. The total available bandwidth in the unlicensed and licensed spectrum is 5 MHz and 10 MHz, respectively. The total power constraint per node is 10 and 20 dBm, which is divided into two parts with the power ratio being fixed at $\alpha = 0.1$. The channel model is constituted by Rayleigh fading and pathloss components, and the pathloss factor is set to $\beta = 3$. The payloads are generated by truncated Pareto distribution with parameters $\kappa_1 = 1.5$ and $\kappa_2 = 3$. And the average payload size is 5 kbits. The constraint on the maximum collection period is 1 s.

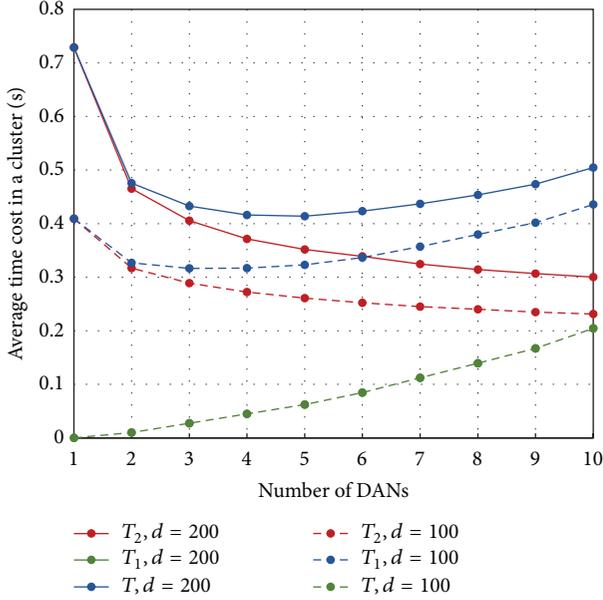


FIGURE 2: The average time cost performance ($\bar{T}_j, \bar{T}_j^*, \bar{T}_j^{(1)}$) versus different number of DANs (M_j) with different cluster locations ($d = 100$ m, 200 m).

Figure 2 shows the variation trend of the average time cost with different number of DANs and cluster locations. It can be seen that, with the increase of the DAN number in one cluster, the average total time cost can be reduced to a minimum value and then begins to increase at a certain position. Specifically, with more DANs, the time cost for cooperative transmission declines correspondingly. However, the declining trend becomes more and more slow. Meanwhile, more DANs create a larger cluster range, which will raise the cost for information sharing. Hence, a certain number of DANs is required for a cluster to obtain the optimal performance in this context. Besides, the cluster location also influences the variation trend of performance. For example, the same clusters at the locations 100 m and 200 m far from the BS have different time cost as well as the switch point of DAN number. The reason is that the farther the distance between cluster and BS is, the more enhancement on performance can be obtained via cooperative transmission, and therefore the larger DAN number is required.

Figure 3 shows the performance comparison between the proposed and the noncooperative schemes with the change of available bandwidth under different conditions of DAN number. Obviously, larger available bandwidth can bring better performance for both schemes. However, the difference between the performance enhancement of these two schemes becomes smaller and smaller. It means that though increasing the bandwidth for a cluster can reduce the time cost, the spectral efficiency will decrease inversely. In addition, Figure 3 also shows that it is reasonable to allocate more frequency resource for clusters with more DANs, since the spectral efficiency in large clusters is higher.

Figure 4 shows the comparison results between the proposed and the noncooperative schemes with the variation

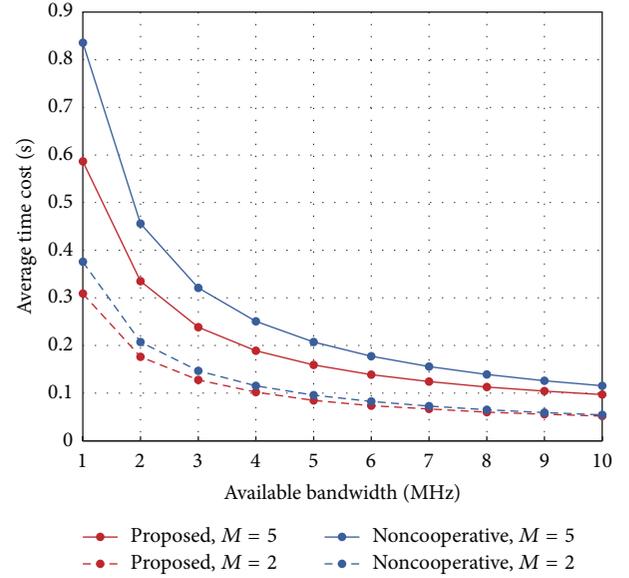


FIGURE 3: Performance comparison between the proposed and the non-cooperative schemes versus different frequency resources (w_j) with different number of DANs ($M_j = 2, 5$).

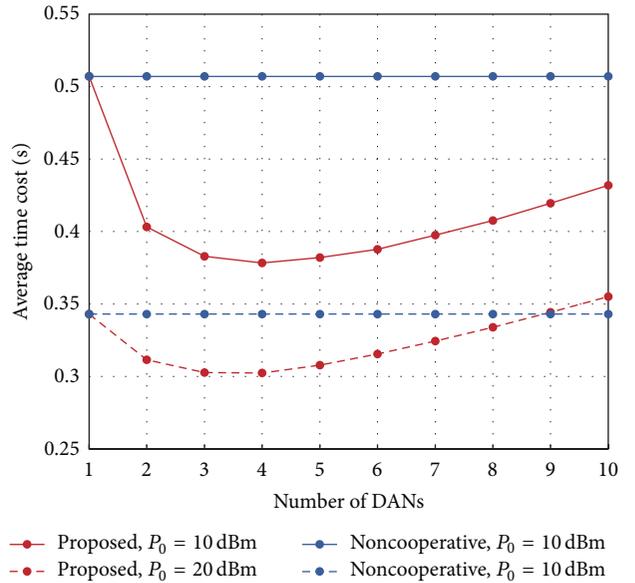


FIGURE 4: Performance comparison between the proposed and the noncooperative schemes versus different number of DANs (M_j) with different transmit power ($P_0 = 10$ dBm, 20 dBm).

of DAN number and transmit power under the condition of the same frequency resources; that is, $w_j = M_j \cdot w_v$ as mentioned in Section 4. It can be seen that, with a certain number of DANs, the performance of the proposed scheme is much higher than the noncooperative scheme due to the advantages of VMIMO. However, if there are too many DANs in a cluster, the performance will decline since the benefit of cooperative transmission is not large enough to cover the cost

for cluster formation and information sharing. Meanwhile, from Figure 4, we can also conclude that more benefit of the proposed scheme can be gotten with lower transmit power. The reason is that the poorer the condition is, the more necessity for DANs to find partner is.

6. Conclusion

In this paper, we focused on the issue of cooperative data collection schemes in M2M networks. To collect the real-time information efficiently, a cluster-based hierarchical transmission scheme with virtual MIMO protocol was provided. In this scheme, the data was firstly shared with intracenter broadcasting. And then all partners transmitted all the total payload cooperatively. To construct the network and allocate resource optimally, we first formulated an optimization problem and derived theoretical analysis results on the advantages and properties of the system. Since the optimal solution can hardly be obtained in practical use, a distributed joint cluster formation and resource allocation scheme was proposed under the context of coalition formation game. Simulation results showed the efficiency of the proposed scheme by evaluating the average time cost performance for data collection.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported in part by the National Science Foundation of China (NFSC, Grant no. 61371073), the National High-Tech Research and Development Program of China (863 Program, no. 2012AA01A506), and the National Science and Technology Major Project under Grant (2013ZX03003003 and 2013ZX03003016).

References

- [1] W. H. Chin, Z. Fan, and R. Haines, "Emerging technologies and research challenges for 5G wireless networks," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 106–112, 2014.
- [2] D. Niyato, L. Xiao, and P. Wang, "Machine-to-machine communications for home energy management system in smart grid," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 53–59, 2011.
- [3] M. J. Booyen, J. S. Gilmore, S. Zeadally, and G. J. van Rooyen, "Machine-to-machine (M2M) communications in vehicular networks," *KSII Transactions on Internet and Information Systems*, vol. 6, no. 2, pp. 529–546, 2012.
- [4] Z. Fan and S. Tan, "M2M communications for e-health: standards, enabling technologies, and research challenges," in *Proceedings of the 6th International Symposium on Medical Information and Communication Technology (ISMICT '12)*, pp. 1–4, March 2012.
- [5] H. Li, Z. Han, L. Lai, R. C. Qiu, and D. Yang, "Efficient and reliable multiple access for advanced metering in future smart grid," in *Proceedings of the IEEE 2nd International Conference on Smart Grid Communications (SmartGridComm '11)*, pp. 440–444, Brussels, Belgium, October 2011.
- [6] R. Yu, Y. Zhang, Y. Chen, C. Huang, Y. Xiao, and M. Guizani, "Distributed rate and admission control in home M2M networks: a non-cooperative game approach," in *Proceedings of the IEEE Conference on Computer Communications Workshops*, pp. 196–200, Shanghai, China, April 2011.
- [7] D. Niyato and P. Wang, "Cooperative transmission for meter data collection in smart grid," *IEEE Communications Magazine*, vol. 50, no. 4, pp. 90–97, 2012.
- [8] 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.
- [9] D. Wu, Y. Cai, L. Zhou, and J. Wang, "A cooperative communication scheme based on coalition formation game in clustered wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 3, pp. 1190–1200, 2012.
- [10] A. Sinha and D. K. Lobiyal, "Performance evaluation of data aggregation for cluster-based wireless sensor network," *Human-Centric Computing and Information Sciences*, vol. 3, no. 1, pp. 1–17, 2013.
- [11] J. Nie, D. Li, Y. Han, and S. Xie, "The optimized method of cluster-head deployment based on GA-WCA in Wireless Sensor Networks," in *Proceedings of the International Conference on Computer Application and System Modeling (ICCSM '10)*, pp. V12449–V12452, Taiyuan, China, October 2010.
- [12] S.-E. Wei, H.-Y. Hsieh, and H.-J. Su, "Joint optimization of cluster formation and power control for interference-limited machine-to-machine communications," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '12)*, pp. 5512–5518, Anaheim, Calif, USA, December 2012.
- [13] T.-C. Juan, S.-E. Wei, and H.-Y. Hsieh, "Data-centric clustering for data gathering in machine-to-machine wireless networks," in *Proceedings of the IEEE International Conference on Communications Workshops (ICC '13)*, pp. 89–94, Budapest, Hungary, June 2013.
- [14] M. C. M. Thein and T. Thein, "An energy efficient cluster-head selection for wireless sensor networks," in *Proceedings of the 1st International Conference on Intelligent Systems, Modelling and Simulation (ISMS '10)*, pp. 287–291, Liverpool, UK, January 2010.
- [15] J. Ding, D. Liu, X. Wang, and H. Wu, "An energy-efficient virtual MIMO transmission scheme for cluster-based wireless sensor networks," in *Proceedings of the IEEE 12th International Conference on Communication Technology (ICCT '10)*, pp. 809–813, Nanjing, China, November 2010.
- [16] X. Wen and D. I. Laurenson, "A resource allocation algorithm for cluster-based cooperative MIMO in wireless sensor networks," in *Proceedings of the 17th International Conference on Software, Telecommunications and Computer Networks (SoftCOM '09)*, pp. 211–215, Hvar, Croatia, September 2009.
- [17] IEEE Standards, "IEEE guide for smart grid interoperability of energy technology and information technology operation with the electric power system (EPS), end-use applications, and loads," *IEEE Standards*, pp. 44–96, 2011.
- [18] X. Feng and C. Leung, "A new optimal transmit and receive diversity scheme," in *Proceedings of the IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM '01)*, pp. 538–541, Victoria, Canada, August 2001.

- [19] Z. Han, D. Niyato, W. Saad, T. Basar, and A. Hjørungnes, *Game Theory in Wireless and Communication Networks-Theory, Models, and Applications*, Cambridge University Press, Chichester, UK, 2012.
- [20] W. Saad, Z. Han, T. Başar, M. Debbah, and A. Hjørungnes, “Physical layer security: coalitional games for distributed cooperation,” in *Proceedings of the 7th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt '09)*, Seoul, Republic of Korea, June 2009.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

