An Autonomous Compensation Game to Facilitate Peer Data Exchange in Crowdsensing

Xiang Yan*, Fan Ye†, Yuanyuan Yang† and Xiaotie Deng*
*Department of Computer Science, Shanghai Jiao Tong University, Shanghai, China, 200240
†Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY 11794, USA

***Abstract*—The rapid penetration of mobile devices has provided ample opportunities for mobile devices to exchange sensing data on a peer basis without any centralized backend. In this paper, we design a peer based data exchanging model, where relay nodes move to certain locations to connect data providers and consumers to facilitate data delivery. Both relays and data providers can gain rewards from consumers who are willing to pay for the data. We first prove the problem of relay node assignment is NP-hard, and provide a centralized optimal method to decide which relay nodes goes to which location with an approximation ratio. Then we define an autonomous compensation game to allow relays make individual decisions without any central authority. We derive a sufficient and necessary condition for the existence of Nash equilibrium. We analyze and compare this distributed game to the centralized social optimal solution, and show that the game incurs small bounded social costs, and efficient under various network sizes, numbers of providers, consumers, and device mobility.**

I. Introduction

The penetration of mobile devices with various sensors has made peer data exchange feasible and valuable in many daily life scenarios. In many places (e.g., roads, parks, airports) there are high densities of mobile devices like smartphones carried by users. Each device can collect sensing data of certain types around its location, and such data may carry important information needed by other users. For instance, a passenger on a bus passing by an accident scene can take a photo, which is important for drivers in nearby blocks so they can know what is causing the jam and how to change the route. Before taking her baby out for a walk, a mother wants to know the air quality distribution around a neighborhood, and pedestrians in the neighborhood could provide such kind of data. In these scenarios, data exchange among peer users provide valuable information, and users are willing to pay a reward to obtain desired data.

In such peer data exchange, the one that possessing certain data is called a *provider* and the one needing data from others is called a *consumer*. Usually, the devices have limited radio transmission range, and the density of mobile devices may not always be high enough to ensure direct connectivity among all consumers and providers. To facilitate data exchange, some *relay nodes*, motivated by the economic incentive, may move to certain *relay locations* to connect consumers and providers, and forward data possibly over multiple hops.

In this paper, we study the following problems: what is the optimal strategy to decide which relay node should go to which location? If consumers requesting the same data can pool their rewards, what is the optimal payment allocation algorithm among relays and providers? The strategy and algorithm must be efficient to incur small overheads in computation and node movements, and effective to incentivize relays and providers for peer data exchange.

978-1-5386-2704-4/17/$31.00 ©2017 IEEE
This work is supported in part by US NSF CSR-1513719, and the National Natural Science Foundation of China (Grant No. 11426026, 61632017, 61173011) and a Project 985 grant of Shanghai Jiao Tong University.

There are several challenges to these questions. First, given multiple relay nodes and relay locations, there is a combinatorial thus exponential space of who goes where, leading to an NP-hard problem. We have to design an efficient algorithm. Secondly, the reward pool must be allocated and paid to relays and providers in a way to compensate them fairly based on their contributions. Third, although a centralized algorithm that computes and dictates which relay goes where may achieve global optimal efficiency, in reality no central authority exists and each relay may make its own decision. The payment must be allocated effectively among peers.

We summarize our contributions to address these challenges:

- We formulate the problem of finding relay locations for relay nodes as an optimization problem in graph theory. We show its NP-hardness and propose a centralized approximation algorithm that decides their payments based on the assignment.
- We design an autonomous compensation game for relay nodes to make decisions of where to go individually. We derive the condition for the existence of pure Nash equilibrium, and provide distributed mechanisms for relay nodes to choose the strategy.
- We show that the social cost, quantified by moving distances of relay nodes, is linear to network size for centralized mechanism. Compared to the social optimal assignment by a central authority, the cost for distributed mechanism is bounded by two measures: Price of Anarchy and Price of Stability. We further analyze how practical factors affect these bounds, and demonstrate our approach remains feasible under provider/consumer mobility.

II. Related Work

Current studies for data exchange among peers focus on two main aspects: efficient and accurate method for managing the exchanging process and fair payment rules for attracting participants.

Helgeson et al. [4] provide a bottom implementation approach for managing data exchange in a network from industrial point of view, and it could be the basis for all data exchange model. Rahman et al. [10] study how to assure security in the exchanging process. More recent work designs data exchange models dealing with further constraints such as ability to recover document damage [8], and adaption to big data [1]. Compared with our approach, they all consider a centralized system to supervise the whole process.

Meanwhile, algorithmic game theory has been partly used to study participants' incentive to exchange data. Gao et al. [3] propose an auction policy to attract more long-term user participation. Luo et al. [7] also design mechanisms for participatory sensing systems to incentivize contribution from users. Similar as our work, they are trying to attract more users to participate through providing rewards, but their designs all require a centralized system to deal with payment.

On the other hand, data transmissions under mobility have raised much scientific interest in the past ten years. Zhao et al [12] started a frame work of controlling the mobility for data transport ferries in a delay-tolerant network, where data is relayed by ferrying nodes and stationary nodes together. After them, Kavitha and Altman [5] study similar message ferry routes design in sensor networks using polling models. Our work considers a different form for delivering data, mainly taking advantages of mobile devices's ability. If the desired data requires huge storage, these data ferry schemes may also be a good choice.

III. Network Model and Basic Assumptions for Data Exchange

In this section, we first present our crowdsensing model for peer data exchange and basic assumptions. Then we illustrate how it works within the whole process through an example.

A. A Crowdsensing Approach for Data Exchange

Our data exchange system consists of several kinds of participants. Besides consumers, providers and relay nodes introduced before, there are some clearance nodes accepting relay information and deciding payments among others.

Before describing the whole process, we first list several important assumptions. Firstly, providers, consumers and clearance nodes are uniformly distributed and their positions do not change in a short time. Clearance nodes can communicate with each other, so each can acquire all the relay information. Each node u entering the system has an unique l-bit account ID number, denoted by A_u and known by themselves and the clearance nodes. Secondly we do not consider transmission failures or cheating nodes. The relay information has much smaller size compared to data, thus their transmission costs are negligible and not considered. Lastly, a relay node transmits one unit of data one time and the amount of transmission is the number of data units.

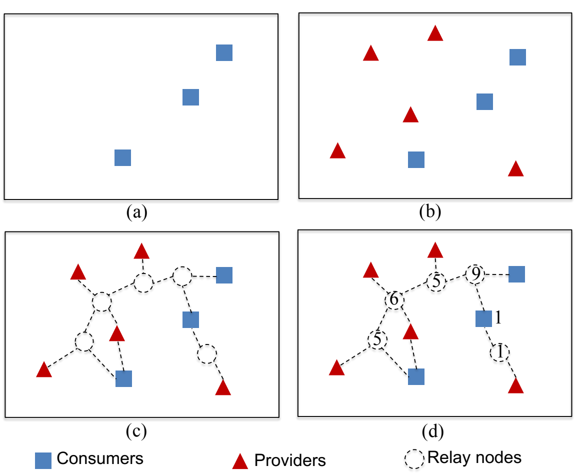

Fig. 1. A complete process for transmitting a group of data. The numbers beside nodes is the corresponding rewards if a relay nodes works there.

With these assumptions, the data exchange system works as shown in Fig. 1. Firstly, consumers send requests for some data, each naming a payment. Then, providers with required data are found. After that, relay locations to connect consumers and providers are found, as well as the delivery paths. Then a central entity calculates the reward for each location, and assigns relay nodes to those locations to finish data delivery. Alternatively, if each relay knows the information about those locations and rewards, it can make its own decision to go to specific location, without the central entity. Once the delivery is done, all involved nodes send messages to clearance nodes, which decide the payments for all relays and providers.

B. Payment Decision for Each Participant

Two key factors decide payments for relays and providers. First, the system should get information for all the actual transmissions. Second, the payments should be "fair" to all participants. We will introduce a transmission graph structure as a record of how a specific piece of data is delivered, then show how payments are decided to achieve fairness.

To construct the transmission graph, providers and relays send short messages to their nearest clearance nodes each time they receive a piece of data from others (i.e., inflow message, denoted by $T^{(i)}$), or send one to the next node (i.e., outflow message, denoted by $T^{(o)}$). Specifically, when data with label L is transmitted from node u to node v, both u and v will send a $3l+|L|$-bit tuple (L, A_g, A_u, A_v), where node g is the original generator of the data and $|L|$ is the length of the label.

After data delivery, all clearance nodes communicate with each other so that all messages about the same data are handled by the same clearance node. For each group of providers and consumers (labeled by $\hat{L}$), such messages are used to build the whole transmission graph $TG(V, E, \hat{g}, \beta)$ by Algorithm 1 (Table I). The algorithm analyzes all messages to obtain participating peers as vertices and corresponding transmissions as edges. $d_{in}(u)$ and $d_{out}(u)$ denote the in/out degrees for vertex u, and they are further used to decide which role the vertex plays in the whole transmission process.

TABLE I
Algorithm 1: Rebuilding Transmission Graph

Input: $\{T_j^{(*)}\}_{j=1}^{\Lambda} = \{(\hat{L}, A_{\hat{g}}, A_{u_j}, A_{v_j})\}_{j=1}^{\Lambda}$.
Output: $TG(\hat{L}) = (V, E, \beta, \hat{g})$.

1: **Set** $V = \bigcup_{j=1}^{\Lambda} \{g_j, u_j, v_j\}$ and $E = \emptyset$
2: **For** $(\hat{u}, \hat{v} \in V)$
  If $\exists j_1, j_2$ such that $u_{j_1} = u_{j_2} = \hat{u}, v_{j_1} = v_{j_2} = \hat{v}$
   Add directed edge $(\hat{u}, \hat{v})$ into E.
3: **For** $(\widetilde{u} \in V)$
  If $d_{in}(\widetilde{u}) > d_{out}(\widetilde{u})$
   $\beta(\widetilde{u}) = 1$
  Else If $d_{in}(\widetilde{u}) < d_{out}(\widetilde{u})$
   $\beta(\widetilde{u}) = 2$
  Else
   $\beta(\widetilde{u}) = 3$

Participants are classified by function $\beta : V \to \{1, 2, 3\}$ into 3 types for payment decision. To be precise, consumers are of type-1 and providers are of type-2. And a relay node receives a piece of data once, but may transmit it to multiple peers. Those having the same out degree and in degree are of type-3, while those having more out degree than in degree are also type-2. Now we present the payment rule.

- Every u with $\beta(u) = 1$, either a consumer or not, pays a price of data (decided in advance), consisting the whole reward pool. This is because vertices of type-1, even though not consumers at the beginning, act as consumers.
- P_1 percent of reward is given to the generator $\hat{g}$. This corresponds to a fairness principle that one producing widely popular data should be better rewarded.
- P_2 percent of reward is given to those of type-2 and 3, and what u receives is proportional to the ratio $\frac{d_{out}(u)}{\sum_{\beta(v) \neq 1} d_{out}(v)}$.

This is awarding each relay node according to the proportion of transmissions it does in all transmissions.

- The rest reward is given to those of type-2. How much u receives is proportional to the ratio $\frac{d_{out}(u)-d_{in}(u)}{\sum\limits_{\beta(v)=2} d_{out}(v)-d_{in}(v)}$. This corresponds to the situation that some relay nodes receive data from one source and transmit it to multiple destinations. They actually makes extra efforts besides data transmission, e.g. storing the data for a long time. And we use the number of transmissions for sending out data minus the number of transmissions for receiving data to measure their extra effort, and this rule awards them according to the proportion of their extra efforts in all such efforts.

IV. Centralized Method to Assign Relay Nodes

In this section, we provide a centralized method to assign relay nodes to relay locations. Here we suppose that the central entity knows all nodes' positions. Providers and consumers do not move before data delivery completes, and relay nodes, supposed to be enough, will follow the assignments. Meanwhile, the energy cost of transmitting each data unit is the same. Then the assignment has two steps. The system first finds necessary locations requiring relay nodes to deliver data. Then it assigns proper potential relay nodes to these locations.

Step 1: Given N_1 providers $\{u_j\}_{j=1}^{N_1}$, N_2 consumers $\{v_j\}_{j=1}^{N_2}$, and transmission range d, the system seeks the least k locations $\{w_j\}_{j=1}^{k}$ and corresponding edges, such that $\forall i, j$ there exists a path from u_i to v_j via some $w_l's$ or $v'_{j'}s$, with all edges shorter than d. We call this *relay location decision problem*. Here we want to minimize the number of relay locations, i.e. the number of necessary relay nodes, so that a centralized system can assign the least number of relay nodes to relay locations.

To show the complexity of relay location decision problem, we refer to a *NP-hard* problem, namely *Steiner tree problem with minimum number of Steiner points* (STP-MSP in short), which is to find the least number of extra points and corresponding edges within bounded length to connect a set of terminal points in two-dimensional Euclidean plane [2]. To be specific, relay nodes problem requires a similar solution with respect to two sets of terminal points, and one of them must be leaf-points, i.e. having degree one.

We reduce the STP-MSP problem to relay location decision problem to show the complexity. The solution to STP-MSP problem must be a tree and at least one of terminal points is leaf-point (because if an extra point is leaf-point, it could be deleted to decrease the number). So solving a STP-MSP problem for $\{p_1, p_2, \ldots, p_n\}$ is exactly solving the relay location decision problem for a p_j being provider u_1 and the rest of points being consumers $\{v_1, v_2, \ldots, v_{n-1}\}$. Thus if we have an algorithm to solve relay location decision problem, we could solve STP-MSP problem by running it n times, selecting a p_j each time as u_1, and find the best result among them. This shows the relay location decision problem is also NP-hard.

Despite its *NP-hardness*, we provide an approximation algorithm to solve it in Table II. As for the approximation ratio, in fact Algorithm 2 without the last loop (line 8) is a ratio-3 approximation algorithm solving STP-MSP problem, and the last loop will increase the number of vertices in the solution by at most N_1. As a result, Algorithm 2 outputs a solution that minimizes the number of relay locations at no more than $3k + N_1$, where k is the number of relay locations in a solution to relay location decision problem. Roughly speaking, the approximation ratio is $N_1 + 3$.

For time complexity, the key part is the third loop (line 6) consisting of at most $O(N_1+N_2)$ rounds (corresponding to the $O(N_1+N_2)$ connected components with at most $O(N_1+N_2)$ vertices each), each of which requires at most $O(N_1+N_2)^3$ operations. Since other parts requires less than $O(N_1+N_2)^3$ operations, the time complexity is $O((N_1+N_2)^4)$.

TABLE II

Algorithm 2: Deciding relay locations

Input: $u_1, u_2, \ldots, u_{N_1}, v_1, v_2, \ldots, v_{N_2}, d$.
Output: Graph $T = (V, E)$.

1: **Set** $V = \{u_1, u_2, \ldots, u_{N_1}, v_1, v_2, \ldots, v_{N_2}\}$, $E = \emptyset$;
2: **Let** e_{uv} be the edge between $u \neq v \in V$;
3: **For** $(u \in \{u_1, u_2, \ldots, u_{N_1}\}, v \in \{v_1, v_2, \ldots, v_{N_2}\})$
  If $e_{uv} \leq d$
    Add e_{uv} into E
4: **Sort** all e_{uv}'s to $\{e_i\}$ in length increasing order;
5: **For** $(i \wedge e_i \leq d)$
  If e_i connects two different connected components of T
    Add e_i into E;
6: **While** (There are more than two connected components) **Do**
  For $(a, b, c \in V$ in three connected components of $T)$
    If $\exists s$, s.t. edge e_{sa}, e_{sb}, e_{sc} shorter than d
      Add s into V and e_{sa}, e_{sb}, e_{sc} into E;
  End while
7: **For** (i)
  If e_i connects two different connected components of T
    Divide e_i into $\lceil \frac{|e_i|}{d} \rceil$ parts and add into T;
8: **For** $(1 \leq i \leq N_1)$
  If u_i has degree larger than one
    Replace u_i by a new vertex u'_i and add (u_i, u'_i) into E

In the output graph, vertices who are not from input are relay locations, in each of which at least one relay node is needed to help transmit data. The reward $r_i, i = 1, 2, \ldots, k$ associated with each of them can be calculated as follows. Firstly, find a path from each provider to each consumer. Then change these undirected edges into directed ones according to the direction of paths they are in to get the transmission graph. Finally use the payment rule introduced in Section III.

Step 2: Given relay locations, the system chooses relay nodes closest to these locations from all potential ones, in the sense that the summation of distances all relay nodes should move is minimized. To obtain this, we can use the *Kuhn-Munkres* algorithm [9], which finds the minimum matching for weighted bipartite graph. We can construct a complete weighted bipartite graph $G(V_1, V_2, E)$ where the relay locations belong to V_1 and the original positions of potential relay nodes belong to V_2. There is an edge between each vertex $u \in V_1$ and each one $v \in V_2$ weighted by their distance. Then we run *Kuhn-Munkres* algorithm to find its minimum matching, which is exactly the optimal assignment for relay nodes.

V. An Autonomous Compensation Game for Individual Relay Node Decisions

The centralized method does not consider the preferences of relay nodes. As shown in Fig. 1. (d), the reward associated with each relay location varies a lot. A centralized system assigns the closest potential relay node to each location, in order to minimize their moving distance so data transmission can begin quickly. However, if relays can make their own decisions, they may choose locations with high rewards, not closest.

We want to design a game where relay nodes make individual decisions of destinations, and we hope there exists Nash equilibrium. Intuitively, a Nash equilibrium is a group of players' strategies where no one could strictly benefit by changing only his own strategy. Specific to our problem, the equilibrium means each relay node goes to a location where it can gain the most reward if others do not change their locations. With the assumption that every player is rational to choose the best strategy for himself, relay nodes will choose locations corresponding to the Nash equilibrium individually.

Now we introduce the autonomous compensation game. We suppose that the optimal positions for relay nodes, as well as the corresponding rewards, are known to all potential relay nodes, either calculated by themselves or broadcast by the system. In this game, the reward of each location will be equally shared by all players moving there. So players need to find a strategy to decide where to move to gain higher utility. Generally speaking, the utility of each player should be the reward he receives by acting as relay minus the energy cost of doing so and other cost for moving to the specific position. The energy cost is proportional to the number of transmission one does, thus proportional to the reward. The moving cost is quite complicated, and we only introduce a simple version due to space limitation, leaving other analysis in a future journal version.

Suppose there are M players ($\{\xi_j\}$) and k locations ($\{w_i\}$) with value (r_i). Define $\mathcal{X} : \{w_i\} \rightarrow 2^{\{\xi_j\}}, \mathcal{X}(w_i) = \{j | \textit{node } \xi_j \textit{ goes to location } w_i\}$. The utility of each player is $U(\xi_j) = \frac{r_i}{|\mathcal{X}(w_i)|}$ where $j \in \mathcal{X}(w_i)$. Here we consider players choose pure strategy, that is each one decides to go to only one location to do transmission in a short period. Then the corresponding pure Nash equilibrium requires

$$\mathcal{X}(w_i) \neq \emptyset, \qquad \forall i \tag{1}$$

$$\frac{r_i}{|\mathcal{X}(w_i)|} \geq \frac{r_j}{|\mathcal{X}(w_j)+1|}, \qquad \forall i \neq j \tag{2}$$

Constraint (1) guarantees that there is at least one relay node at each relay location, so the network is connected. Constraint (2) ensures that when everyone chooses the strategy associated to the equilibrium, each player cannot gain more utility by changing his strategy when others keep their strategies. Here players are homogeneous since they choose the strategies to maximize their utilities without considering the cost generated during their moving to specific locations. As a result, to find an equilibrium, we only need to study the number of players going to each location.

Theorem 1. *A sufficient and necessary condition for the existence of pure Nash equilibrium is:*

$$M \geq \sum_{i=1}^{k} \max\{1, \lceil \frac{r_i}{r_k} \rceil - 1\} \tag{3}$$

where r_i is in decreasing order. And the equilibrium is unique if the equality holds.

We give a sketch of proof and leave the complete one in a future journal version. From constraint (1), at least one peer occupies location w_k. From his point of view, if there exists a location w_t with reward r_t satisfying: $r_t > m \cdot r_k$ but $|\mathcal{X}(w_t)| < m$, then he will go to w_t to improve his utility. This means the reward of location w_i should be divided by at least $\lceil \frac{r_i}{r_k} \rceil - 1$ agents once $\frac{r_i}{r_k} > 2$. On the other hand, if the number of peers are more than $M_0 = \sum_{i=1}^{k} \max\{1, \lceil \frac{r_i}{r_k} \rceil - 1\}$, we can first construct an equilibrium allocation for M_0 agents, and prove there is an equilibrium for each $M > M_0$ by induction. Finally the uniqueness is proved by contradiction.

Remark: In practice, we also need to define d_{ij} as the distance between w_i and ξ_j, and assume the cost for ξ_j moving to w_i is $c \cdot d_{ij}$ with a constant factor c. Then the Nash equilibrium requires for $\forall i$ and $\forall j \in \mathcal{X}(w_i)$

$$\mathcal{X}(w_i) \neq \emptyset \tag{4}$$

$$\frac{r_i}{|\mathcal{X}(w_i)|} - c \cdot d_{ij} > 0 \tag{5}$$

$$\frac{r_i}{|\mathcal{X}(w_i)|} - c \cdot d_{ij} \geq \frac{r_{i'}}{|\mathcal{X}(w_{i'})+1|} - c \cdot d_{i'j}, \quad \forall i' \neq i \tag{6}$$

Generally speaking, finding the Nash equilibrium is solving a system of integer equations, which is an *NP-hard* problem and there is no common method for approximate solutions. With similar analysis as the one before, we can still have a necessary condition for the existence of Nash equilibrium

$$M \geq \sum_{i=1}^{k} \max\{1, \lceil \frac{r_i}{r_k + c \cdot (d_{max} - d_{min})} \rceil - 1\} \tag{7}$$

where $d_{max} = \max\{d_{ij}, 1 \leq i \leq k, 1 \leq j \leq M\}$ and $d_{min} = \min\{d_{ij}, 1 \leq i \leq k, 1 \leq j \leq M\}$.

Although the pure Nash equilibrium is assured if there are relative enough players participating, in most circumstance, the peers join the game, meaning deciding to relay data, in a series of time. Assuming player ξ_1 joins firstly, followed by ξ_2, and so on, we show how the equilibrium is approached: (1) Before the M_0-th player joins, each player chooses the location with highest actual reward, that is the reward for each peer going there, at the moment he decides to participate. If there are multiple such locations, choose the one where no relay node goes yet to avoid a failure of whole system, or randomly one if these locations are occupied by at least one relay node. (2) After that, each player still chooses the location with highest actual reward when he comes. But if there are multiple such locations, choose the one with highest total reward. According to the proof of Theorem 1, we know the equilibrium always holds after any player joins in this way.

VI. Performance Evaluation for Data Exchange System

In this section, we evaluate our design in three aspects. We first show that the transmission overhead is roughly a linear function of the network size, or numbers of providers/consumers. Then we show how practical provider/consumer mobility impacts the performance. Finally, we analyze Nash equilibrium and find the upper/lower bounds of the extra costs incurred when potential relay nodes make individual decisions.

A. Transmission Overhead as Functions of Network Size

We present the relationship between transmission overhead and the numbers of providers and consumers in the network. Since there is no closed form solution for finding proper positions for relay nodes, we use approximation algorithms.

Fig. 2 shows the total transmission overhead as functions to the numbers of providers/consumers, and network diameter. The results are averaged over 1000 runs, where all consumers, providers and potential relay nodes are independently and uniformly distributed in a $D \times D$ square. We can see that when the transmission overhead increase linearly as the number of providers or consumers increases when either remains fixed.

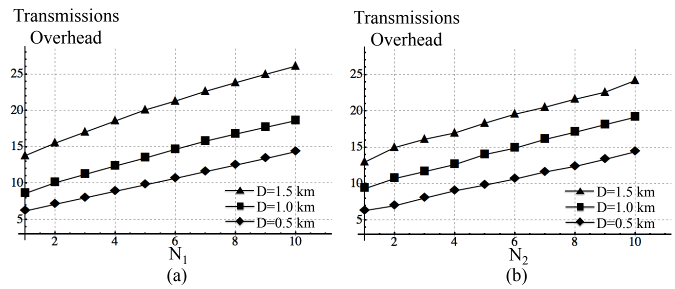

Fig. 2. Total transmission overhead as functions of numbers of providers/consumers (N_1/N_2) with different D's.

B. Impact of Practical Mobility

In reality, providers and consumers are not always static. To analyze how devices' mobility influences data exchange, we follow the model in [6] that finds the nodes' speeds and pause times each follow a log-normal distribution.

We assume providers and consumers are walking around slowly in the $D \times D$ square, and following Log-Normal distributions denoted by $ln\mathcal{N}(\mu, \sigma^2)$. Potential relay nodes can move at $10km/h$ on average to assigned locations, and providers/consumers will stop once all relay nodes reach their destinations and data delivery starts. Here we define a concept of *second move*. When relay nodes reach their destinations, some of these locations may no longer be suitable for data delivery because providers/consumers have moved. Thus a fraction of relay nodes may need to move a second time to some new locations.

Fig. 3 shows how the average speed of providers/consumers impact second move overhead. It varies from static, normal walking (3km/h) to slow running (5km/h). The probability that second move happens increases almost linearly as providers/consumers move faster (Fig. 3.(a)). This is intuitive because the faster they move, the more likely some original relay locations become obsolete. The fraction of relay nodes that need a second move, however, fluctuates but remains at a low percentage ($8 \sim 12\%$ in Fig. 3.(b)). Thus when second move is needed, only one in ten relay node is affected. Also the total second move distance as a ratio of previous move's total distance, is also small ($7 \sim 15\%$ in Fig. 3.(b)). These show that the mobility of providers/consumers has small chance ($5 \sim 25\%$) of incurring a very small additional overhead.

C. Metric for Nash Equilibrium

In autonomous compensation game, each potential relay node chooses strategy of going to some relay location individually. It is natural that more relay nodes go to those relatively "richer" locations in the equilibrium. They work together to relay data, even though one is enough, which is the "inefficiency" due to individual choices. We use a welfare function of cost form to measure the inefficiency, and compare it to the social optimal assignment. Since more than one Nash equilibrium may exist, the best or the worst equilibrium as quantified by the welfare correspond to the two metrics of Price of Anarchy (*PoA* in short) and Price of Stability (*PoS*).

1) Price of Anarchy: Let S be the joint strategy sets for players $E \subseteq S$ be the set of strategies in equilibrium. In this game, the total moving cost $C : S \rightarrow \mathbb{R}$ for a group of players' strategies is considered as social welfare which should be minimized, and the optimal solution can be found according to the analyze in Section IV. Then Price of Anarchy is defined as the ratio between the social welfare of the "worst equilibrium" and the one of optimal "centralized" solution. In the simplest model, all relay nodes ignore the moving cost when choosing strategies. With previous notations, a lower bound for optimal social welfare is $k \cdot d_{min}$ and an upper bound for the worst equilibrium is $M \cdot d_{max}$. So the correspondingly upper bound for Price of Anarchy is

$$PoA = \frac{\max_{s \in E} C(s)}{\min_{s \in S} C(s)} \leq \frac{M \cdot d_{max}}{k \cdot d_{min}}$$

Intuitively, when more people are doing a fixed amount of work, there will be more unnecessary cost. And when more amount of work are given to a fixed number of people, there will be less waste due to competition.

2) Price of Stability: With the same setting as *PoA*, *PoS* measures the ratio between the "best equilibrium" and the optimal "centralized" solution. By definition, $1 \leq PoS \leq PoA$. The closer they are to 1, the less inefficiency in the equilibrium. The metric of *PoA* is an upper bound, and the metric of *PoS* is the lower bound for the inefficiency in an equilibrium. Here the optimal solution is exactly the same as before. For the best equilibrium, the upper bound corresponds to the assignment where k nodes go to the same locations as in the optimal solution, and others go to furthest locations. Then we deduce an upper bound for *PoS*:

$$PoS = \frac{\min_{s \in E} C(s)}{\min_{s \in S} C(s)} \leq 1 + \frac{(M-k) \cdot d_{max}}{k \cdot d_{min}}$$

D. Numeric Evaluation of PoA/PoS

We evaluate the impact of different factors in PoA/PoS following similar settings with simulation.

Fig. 4. shows that PoA and PoS are monotonously decreasing as the number of providers (consumers) increases while number of consumers (providers respectively) fixed. The curves are similar to inverse proportional functions, consistent with our earlier theoretical analysis.

Fig. 5 shows PoA and PoS are both linear functions of the number of potential relay nodes (M), similar as previous analysis. We also find that the gap between *PoS* and *PoA* becomes narrower when more providers/consumers exist. The gap is caused by multiple Nash equilibriums. Thus less equilibrium and narrower gap exist when there are more providers and consumers.

Fig. 6 shows that *PoA* and *PoS* follow earlier analysis at when network diameter (D) is small but becomes unreasonable when $D > 1.5km$. This may be caused by the fixed price for each piece of data, regardless to the distribution of providers and consumers. When peers are in a larger area, more relay nodes are needed, resulting in a decrease of reward to each of them. This shows increasing the price for data for larger network diameter is necessary.

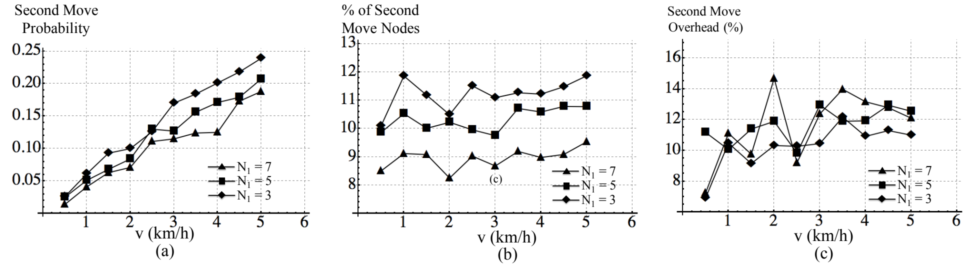

Fig. 3. Performance of our approach as a function of average speed v (km/h) of providers and consumers with impact of N_1 (N_2 fixed). (a) Second move probability vs v. (b) Fraction of second move nodes vs v. (c) Second move overhead vs v.

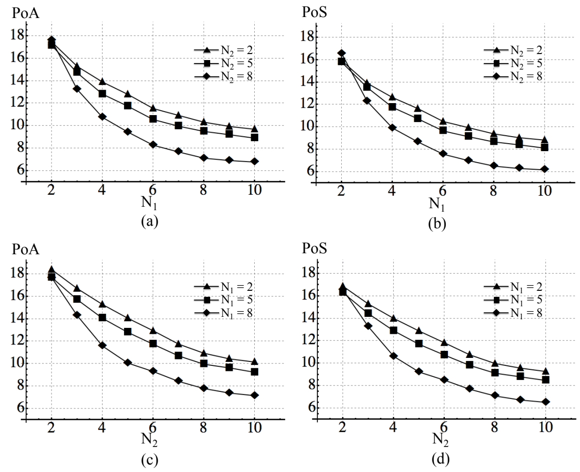

Fig. 4. *PoA/PoS* as a function of N_1 (N_2) with impact of N_2 ($N_1 respectively$). $M = 50$, $D = 1km$.

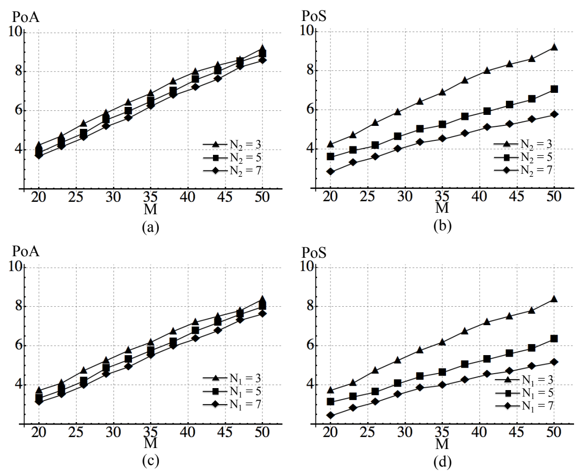

Fig. 5. *PoA/PoS* as a function of M with impact of N_2 (N_1 fixed) or N_1 (N_2 fixed respectively). $D = 1km$. (a) *PoA* vs M when $N_1 = 3$. (b) *PoS* vs M when $N_1 = 3$. (c) *PoA* vs M when $N_2 = 3$. (d) *PoS* vs M when $N_2 = 3$.

VII. Conclusions

We design a crowd sensing incentive framework for peer data exchange where consumers need data from providers, and relay nodes facilitate the exchange by going to relay locations to connect them. Relays and providers gain utilities by relaying or generating desired data. We first propose a centralized method to assign relay nodes to locations, then introduce a new autonomous compensation game model for them to make decisions individually. We analyze the condition for the existence of the Nash equilibrium, and evaluate how different factors impact the overhead. We also compare the inefficiency in individual decision to that of social optimal assignment, and find that it does not increase too much when participants have more freedom choosing their strategies, due to the bounds in

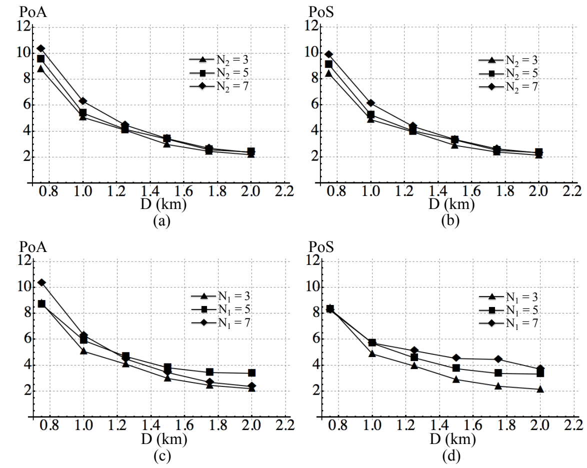

Fig. 6. *PoA/PoS* as a function of D with impact of N_2 (N_1 fixed) or N_1 (N_2 fixed respectively). $M = 100$. (a) *PoA* vs D when $N_1 = 3$. (b) *PoS* vs D when $N_1 = 3$. (c) *PoA* vs D when $N_2 = 7$. (d) *PoS* vs D when $N_2 = 7$.

the increasing rate of the Price of Anarchy/Stability.

References

[1] Asad Z, Chaudhry M A R, Malone D. Greener Data Exchange in the Cloud: A Coding-Based Optimization for Big Data Processing[J]. IEEE Journal on Selected Areas in Communications, 2016, 34(5): 1360-1377.

[2] Du D, Wang L, Xu B. The Euclidean bottleneck Steiner tree and Steiner tree with minimum number of Steiner points[M]//Computing and Combinatorics. Springer Berlin Heidelberg, 2001: 509-518.

[3] Gao L, Hou F, Huang J. Providing long-term participation incentive in participatory sensing[C]//2015 IEEE Conference on Computer Communications (INFOCOM). IEEE, 2015: 2803-2811.

[4] Helgeson C S, Lipkin D S, Larson R S, et al. Method and apparatus for managing data exchange among systems in a network: U.S. Patent 6,643,652[P]. 2003-11-4.

[5] Kavitha V, Altman E. Analysis and design of message ferry routes in sensor networks using polling models[C]//WiOpt'10: Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks. 2010: 62-70.

[6] Kim M, Kotz D, Kim S. Extracting a Mobility Model from Real User Traces[C]//INFOCOM. 2006, 6: 1-13.

[7] Luo T, Tan H P, Xia L. Profit-maximizing incentive for participatory sensing[C]//IEEE INFOCOM 2014-IEEE Conference on Computer Communications. IEEE, 2014: 127-135.

[8] Milosavljevic N, Pawar S, El Rouayheb S, et al. Efficient algorithms for the data exchange problem[J]. IEEE Transactions on Information Theory, 2016, 62(4): 1878-1896.

[9] Munkres J. Algorithms for the assignment and transportation problems[J]. Journal of the society for industrial and applied mathematics, 1957, 5(1): 32-38.

[10] Rahman S M M, Masud M M, Hossain M A, et al. Privacy preserving secure data exchange in mobile P2P cloud healthcare environment[J]. Peer-to-Peer Networking and Applications, 2015: 1-16.

[11] Wennerström H, Smith D B. A game theoretic approach to sensor data communications in an opportunistic network[C]//2015 IEEE International Conference on Communications (ICC). IEEE, 2015: 6306-6311.

[12] Zhao W, Ammar M, Zegura E. Controlling the mobility of multiple data transport ferries in a delay-tolerant network[C]//Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE, 2005, 2: 1407-1418.

[13] Zhang X, Xue G, Yu R, et al. Truthful incentive mechanisms for crowdsourcing[C]//2015 IEEE Conference on Computer Communications (INFOCOM). IEEE, 2015: 2830-2838.