

# COUPON: A Cooperative Framework for Building Sensing Maps in Mobile Opportunistic Networks

Dong Zhao, Huadong Ma, *Member, IEEE*, Shaojie Tang, *Member, IEEE*, and Xiang-Yang Li, *Senior Member, IEEE*

**Abstract**—Human-carried or vehicle-mounted sensors can be exploited to collect data ubiquitously for building various sensing maps. Most of existing mobile sensing applications consider users reporting and accessing sensing data through the Internet. However, this approach cannot be applied in the scenarios with poor network coverage or expensive network access. Existing data forwarding schemes for mobile opportunistic networks are not sufficient for sensing applications as spatial-temporal correlation among sensory data has not been explored. In order to build sensing maps satisfying specific sensing quality with low delay and energy consumption, we design COUPON, a novel cooperative sensing and data forwarding framework. We first notice that cooperative sensing scheme can eliminate sampling redundancy and hence save energy. Then we design two cooperative forwarding schemes by leveraging data fusion: Epidemic Routing with Fusion (ERF) and Binary Spray-and-Wait with Fusion (BSWF). Different from previous work assuming that all packets are propagated independently, we consider that packets are spatial-temporal correlated in the forwarding process, and derive the dissemination law of correlated packets. Both the theoretic analysis and simulation results show that our cooperative forwarding schemes can achieve better tradeoff between delivery delay and transmission overhead by both by theory and simulation analysis. We also evaluate our proposed framework and schemes with real mobile traces. Extensive simulations demonstrate that the cooperative sensing scheme can reduce the number of samplings by 93% compared with the non-cooperative scheme; ERF can reduce the transmission overhead by 78% compared with Epidemic Routing (ER); BSWF can increase the delivery ratio by 16%, and reduce the delivery delay and transmission overhead by 5% and 32% respectively, compared with Binary Spray-and-Wait (BSW).

**Index Terms**—Mobile opportunistic networks; opportunistic sensing; people-centric sensing; data fusion; routing.

## 1 INTRODUCTION

Recent advancements in mobile sensing and communication technologies, especially the proliferation of smart phones, offers novel, efficient ways to opportunistically collect data, enabling numerous monitoring applications for obtaining sensing maps of air quality [1], noise [2], [3], temperature [4],  $CO_2$  concentration [5], etc. The application scenario is illustrated in Fig. 1: a group of mobile phone users equipped with various sensors, GPS receivers, and wireless communication modules (e.g., bluetooth, Wi-Fi) roam within the monitoring region; each user opportunistically takes samples and reports its sensory data to the monitoring center; the monitoring center then calculates the distribution of sensory data and presents the result on the maps (e.g., Google map). Such sensing paradigm is popularly called *opportunistic sensing* [6], *participatory sensing* [1], [3], *mobile crowd-*

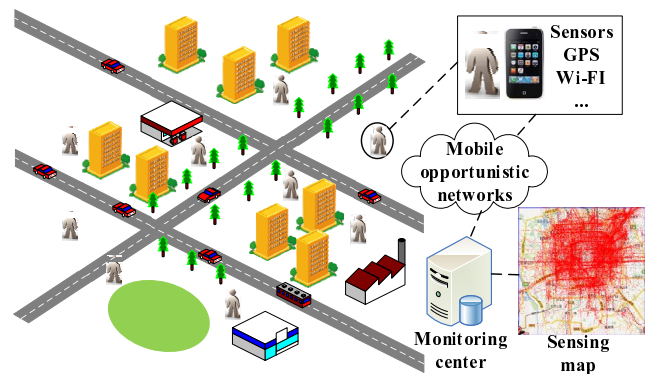


Fig. 1. The illustration of an opportunistic sensing system.

*sourced sensing* [7], [8] or *people/human-centric sensing* [6], [9], [10].

Most of existing mobile sensing applications consider mobile users who can report and access sensory data through the Internet by cellular networks (e.g., 3G/4G mobile networks) or Wi-Fi connections. However, this approach cannot be applied in some scenarios where network coverage is poor or network access is expensive. For example, dead spots of network coverage are commonly found in remote areas and even in some part of major cities [11]. For another, the infrastructure (such as power and cell towers) is down in disaster

- D. Zhao and H. Ma are with the Beijing Key Lab of Intelligent Telecomm. Software and Multimedia, Beijing University of Posts and Telecomm., Beijing, 100876, China. E-mail: {dzhao, mhd}@bupt.edu.cn.
- S. Tang is with the Department of Computer and Information Sciences, Temple University, USA. E-mail: tangshaojie@gmail.com.
- X.-Y. Li is with the Department of Computer Science, Illinois Institute of Technology, USA. E-mail: xli@cs.iit.edu.

recovery scenarios [12]. On the contrary, opportunistic data forwarding among mobile users becomes possible through intermittent connections with short-range radio communications, which offers another way to collect and share data in areas with poor or no network coverage. It works well without requiring any centralized server or infrastructure for communication and management, and also reduces the workload of cellular networks in dense areas. Moreover, the opportunistic data forwarding approach is more energy-efficient and less expensive, which is very important because most users hope to save the battery energy and data usage of their mobile phones. To fully explore these potentials, it is necessary to design new efficient sensing and data forwarding schemes.

Many opportunistic forwarding protocols, such as Epidemic Routing (ER) [13] and its variations, have been proposed to route data among mobile nodes in Mobile Opportunistic Networks (MONs) [14] and Delay Tolerant Networks (DTNs) [15]. Recently, some work proposed to use mobile phone users as data mules for collecting sensory data opportunistically [11], [12], [16], [17]. However, the spatial-temporal correlation among sensory data and its impact on the network performance have not been thoroughly investigated. In this paper, we integrate the opportunistic forwarding protocols with data fusion (or aggregation) for two reasons. First, the users may be interested only in the aggregated results of the sensory data; for example, only the average value of the relevant parameter (e.g. temperature, noise) may be of interest. Secondly, sensory data collected in close proximity or time period may be highly correlated, and data fusion can effectively eliminate redundancy and hence reduce network overhead. Although many routing schemes supporting data fusion have been proposed in WSNs [18], to the best of our knowledge, we are the first to investigate the opportunistic forwarding schemes supporting data fusion in MONs.

Although the idea of integrating opportunistic forwarding protocols with data fusion seems simple and straightforward, we still face some new challenges for both performance modeling and protocol design in practice. Previous work on performance modeling of opportunistic forwarding protocols assumed that all packets are propagated independently. However, the packets are spatial-temporal correlated in the forwarding process with data fusion, which causes a more complex propagation process. We derive an ordinary differential equation (ODE) model for analyzing the dissemination law of correlated packets, and theoretically prove the bounds of transmission overhead and delivery delay. Our analysis framework can serve as fundamental guidelines on integrating various opportunistic forwarding protocols with data fusion, and achieving desirable tradeoff among various performance metrics. On the other hand, Binary Spray-and-Wait (BSW) [19] is considered to be one of the most efficient methods to reduce the large overhead of the ER scheme without incurring significant delay penal-

ties. However, when we consider integrating BSW with data fusion, an important question must be answered: how many forwarding tokens should be assigned to the nodes for the new fused packet? In order to solve this problem, we design a series of rules to assign proper number of forwarding tokens to the nodes, and theoretically prove the performance improvement in terms of transmission overhead and delivery delay.

There are also some other challenges, such as incentive mechanism design [20], [21] and privacy preserving problems [22], [23], that have attracted much attention from researchers recently. But they are out of the scope of our discussion in this paper.

As a summary, the contributions of this paper are:

- We propose COUPON, a novel cooperative sensing and data forwarding framework for building sensing maps while satisfying specific sensing quality with low delay and energy consumption. The cooperative sensing scheme can eliminate sampling redundancy and hence save energy. Through cooperative forwarding schemes with data fusion, we can improve the network performance in terms of delivery ratio, delivery delay and transmission overhead significantly.
- We design two novel protocols by leveraging data fusion: Epidemic Routing with Fusion (ERF) and Binary Spray-and-Wait with Fusion (BSWF). We derive the dissemination law of correlated packets, theoretically prove that ERF has the same delivery delay with ER [13] while significantly reducing the transmission overhead, and BSWF can reduce both the delivery delay and transmission overhead compared with BSW [19].
- Extensive simulations are conducted based on real mobile traces [24], which demonstrate that the cooperative sensing scheme can reduce the number of samplings by 93% compared with the non-cooperative scheme; ERF can reduce the transmission overhead by 78% compared with ER; BSWF can increase the delivery ratio by 16%, and reduce the delivery delay and transmission overhead by 5% and 32% respectively, compared with BSW.

Compared to our previous conference version of this work [25], this paper presents more theoretical analysis and additional simulation results. The remainder of this paper is organized as follows: Section 2 reviews the related work. Section 3 describes the system model and design objective of COUPON. In Section 4, we introduce the framework of COUPON, including the cooperative sensing scheme and the basic idea of the cooperative forwarding with data fusion. We detail the two cooperative forwarding schemes, and analyze the effects of data fusion on network performance in Section 5. In Section 6, we conduct extensive simulations to evaluate our solution based on real mobile traces. Finally, we conclude the paper in Section 7.

## 2 RELATED WORK

### 2.1 People-Centric Sensing

Recently, many people-centric sensing applications have been developed to leverage human-carried or vehicle-mounted sensors to share local data, increase global awareness, compute community statistics, or map physical phenomena. These applications can be classified into three categories: *environment-centered*, *infrastructure and facility related*, and *social or community centered*. In this paper, we mainly focus on the periodic monitoring applications that generally belong to the first two categories. In the first category, the Common Sense project [1] developed an urban sensing system using handheld air quality monitors. NoiseTube [2] and Ear-Phone [3] were two urban noise mapping systems that created open and inexpensive platforms for rendering up-to-date noise maps. A vehicle sensor network was developed to monitor the  $CO_2$  concentration in areas of interest [5]. The compressed sensing approach was proposed to recover the sensing map from random and incomplete samples more effectively [3], [4]. The relationship between the sensing delay of an urban vehicular sensing system and the number of vehicles was analyzed based on taxis mobility traces [26]. In the second category, Sensorly [27] was a sensing platform for constructing cellular/WiFi network coverage maps. CarTel [28] was a mobile sensor computing system designed to collect, process, deliver, and visualize data from sensors located on mobile units such as automobiles, which can provide urban sensing information such as traffic conditions. Pothole Patrol ( $P^2$ ) system [29] used the inherent mobility of the participating vehicles, opportunistically gathering data from vibration and GPS sensors, and processing the data to assess road surface conditions. ParkNet [30] detected available parking spots using ultrasonic sensing devices installed on cars, combined with smart phones.

Most of existing applications exploit the data reporting approach through the Internet by cellular networks or Wi-Fi connections. By contrast, MetroSense [6] is a novel opportunistic sensor networking architecture that leverages mobility-enabled interactions and provides coordination between people-centric mobile sensors, static sensors, and edge wireless access nodes in support of *opportunistic sensing*, *opportunistic tasking*, and *opportunistic data collection*. Some fundamental issues in the opportunistic sensor networking paradigm have been discussed and partly solved, such as networking performance [6], [16], security and privacy [6], [22], [23].

### 2.2 Opportunistic Forwarding Protocols

Many opportunistic forwarding protocols have been proposed in MONs [14] and DTNs [15]. Among these protocols, ER [13] is the first and the most generic one. It has the minimum delivery delay under ideal conditions at the cost of heavy overhead, and thus can serve as a baseline method. In order to reduce the large transmission overhead, many variants of ER have

been proposed, e.g.,  $K$ -hop schemes [31], probabilistic forwarding [32], spray-and-wait [19], etc. More recently, some work proposed to use mobile phone users as data mules for collecting sensory data opportunistically. Ngai et al. [11] presented a context-aware sensing data dissemination framework for mobile phone users in a remote sensing field. Uddin et al. [12] presented a picture delivery service for disaster-response applications where a group of survivors and first responders sent images to a rescue center using mobile phones in the absence of a functional communication infrastructure. Ngai et al. [16] presented a novel design to provide efficient opportunistic information exchange for mobile phone users in sensing field with data repositories that tackled the fundamental availability and overhead issues. Park et al. [17] implemented a mobile-phone-based data muling system, and used it in four sensornet deployments including rural region, residential and office environments. However, these work failed to consider the spatial-temporal correlation among sensory data and its impact on the network performance.

### 2.3 Data Fusion

Many routing schemes supporting data fusion have been exploited in WSNs [18]. They have been proved to be effective strategies to curtail the network load and hence reduce energy consumption by fusing correlated data along gathering routes. However, these work cannot be applied for dynamic and mostly disconnected network such as MONs and DTNs. On the contrary, we design opportunistic forwarding protocols by leveraging data fusion, and verify that network performance can be significantly improved in terms of delivery delay and transmission overhead.

## 3 SYSTEM MODEL

We consider a MON composed of  $N$  mobile nodes  $U = \{u_1, u_2, \dots, u_N\}$  equipped with sensors in a sensing field. Each node has a movement trajectory, which is a sequence of time-ordered GPS points. The position of node  $u_k$  at time  $t$  is denoted by  $L_k(t)$ . Each node has a sampling period with the size of  $T_s$ , and it can take samples at the end of each sampling period. The mobile nodes forward sensory data using short-range radio communications in the “store-carry-forward” way for saving battery energy and data usage. Namely, a node receiving a packet buffers and carries that packet as it moves, and passes the packet to new nodes that it encounters, until the packet is received by one sink node. Here we assume that some sink nodes are deployed in the network, which can connect with the monitoring center directly. Finally, the monitoring center aggregates all the sensory data, and builds a sensing map. Both the sensing and data forwarding processes are opportunistic due to the randomness of node mobility and intermittent connectivity of the network. We aim to build a sensing map while satisfying specific sensing quality with low

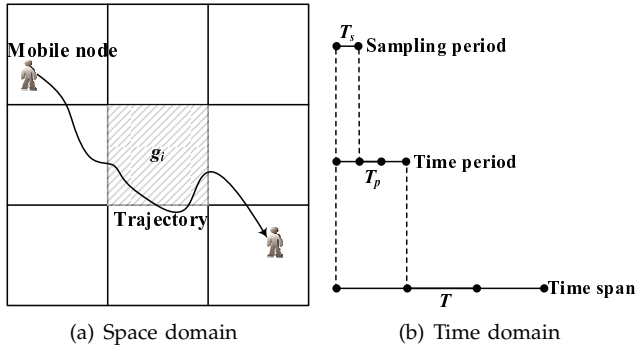


Fig. 2. The illustration of discretizing the time-space domain.

delay and energy consumption. In the following, we introduce these performance metrics respectively.

### 3.1 Sensing Quality

In this paper, we use a basic metric called *coverage* to characterize the quality of a sensing map. The sensory data is often highly correlated in the time and space domains. For example, the air quality at one point in the sensing field at some time could represent the air quality in its neighborhoods within a time period. Thus, we need to discretize the time-space domain of the sensing coverage. In space domain, the whole sensing field is divided into a set of grid cells  $G = \{g_1, g_2, \dots, g_m\}$ , as illustrated in Fig. 2(a). The size of grid cells represents the spatial sensing granularity, which is decided by the application requirements. Assume that we would like to build a sensing map during a time span  $T$ , including  $l$  sampling periods, i.e.,  $T = l \times T_s$ . A grid cell is said to be covered by a node only when a new sampling period arrives and the location of the node is just within the area of the grid cell. Let  $C(g_i, L_k(t))$  denote whether the grid cell  $g_i$  is covered by the node  $v_k$  at the instant  $t$ , i.e.,

$$C(g_i, L_k(t)) = \begin{cases} 1, & \text{if } t \in \{T_s, 2T_s, \dots, lT_s\} \text{ and } L_k(t) \in g_i; \\ 0, & \text{otherwise.} \end{cases}$$

In time domain, we further divide  $T$  into  $l'$  time periods with the same size of  $T_p$ , i.e.,  $T = l' \times T_p$ ,  $T_s < T_p < T$ , as illustrated in Fig. 2(b). Note that we may need a grid cell be covered once within a time period instead of being covered all the time. So the size of time periods represents the required temporal sensing granularity, which is also decided by the application requirements. To build an accurate sensing map, we consider a general definition of coverage constraints:

**Definition 1 (Effective Coverage).** A grid cell  $g_i \in G$  is said to be effectively covered within the  $x$ -th ( $x = 1, 2, \dots, l'$ ) time period, if the number of times that it is covered by all mobile nodes in  $U$  within this time period is not less than  $K$  ( $K \geq 1$ ), i.e.,  $\sum_{t=(x-1)T_p}^{xT_p} \sum_{k=1}^n C_i(L_k(t)) \geq K$ .

Here we define a general coverage constraint,  $K \geq 1$ , to avoid the measurement error and obtain better sensing quality. For example,  $K$  samples of the same grid

cell can be fused (e.g., averaging, summation, voting, max/min, etc.) to obtain more accurate values. A similar concept is the  $K$ -coverage in stationary WSNs, where each point in the target area should be within the sensing range of at least  $K$  sensors, because it provides redundancy and fault tolerance, and facilitate some functions such as detection, localization, classification and tracking of targets [33], [34]. A larger  $K$  provides the grid cells the better sensing quality at the cost of more redundant samplings and energy consumption.

### 3.2 Delivery Delay

**Definition 2 (Delivery Delay).** For an effectively covered grid cell within the  $x$ -th ( $x = 1, 2, \dots, l'$ ) time period, its delivery delay is the duration from the beginning of this time period to the time when  $K$  sensory data sampled in this grid cell within this time period has been delivered to the monitor center.

If the delivery delay is too high, the sensory data will be out of date. Thus, we give a Time to Live (TTL) constraint for each grid cell. Namely, all packets generated within the  $x$ -th ( $x = 1, 2, \dots, l'$ ) time period will be forwarded among mobile nodes only within the duration of TTL from the beginning of this time period.

**Definition 3 (Effective Delivery).** For an effectively covered grid cell within the  $x$ -th ( $x = 1, 2, \dots, l'$ ) time period, it is said to be effectively delivered, if  $K$  sensory data sampled in this grid cell within this time period has been delivered to the monitor center before the TTL expires.

### 3.3 Energy Consumption

The energy consumption of the system is involved in two processes: sensing and data forwarding. The energy consumption in the sensing process is mainly decided by the total number of samples. Without global coverage information or centralized controlling mechanism, mobile nodes will make redundant samples. Thus, a distributed cooperative sensing scheme is necessary to reduce the sampling redundancy by leveraging spatial-temporal correlation among sensory data.

The energy consumption in the data forwarding process depends on the transmission overhead (namely the total number of packets forwarded) of the adopted forwarding scheme. Although many opportunistic forwarding protocols have been proposed, to the best of our knowledge, none of them considered the spatial-temporal correlations among sensory data. In fact, for each effectively covered grid cell, there are at least  $K$  sensory data that are spatial-temporal correlative in the same time period. Thus, a cooperative forwarding scheme will be beneficial for reducing the transmission overhead of the network by fusing the sensory data in the forwarding process.

## 4 COUPON: COOPERATIVE FRAMEWORK

In order to optimize the system performances in terms of sensing quality, delivery delay and energy consumption, we design two cooperative schemes: through cooperative sensing, the specific sensing quality can be achieved with reduced energy consumption; through cooperative forwarding with data fusion, it can obtain higher delivery ratio with reduced delivery delay and transmission overhead.

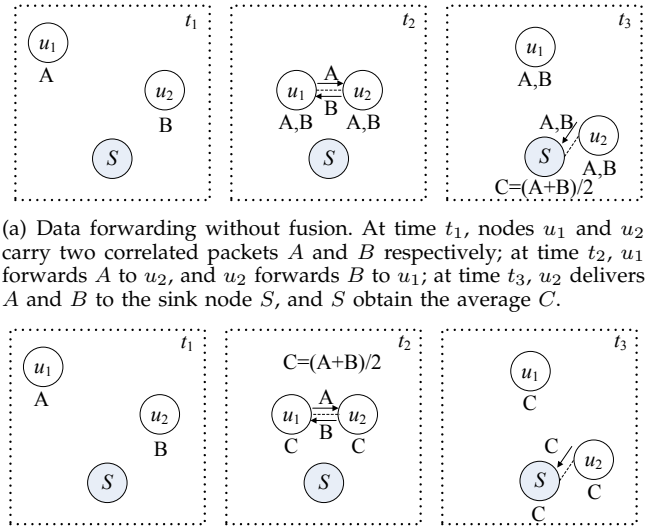
### 4.1 Cooperative Sensing

Without centralized controlling mechanism, each node only locally decides whether to perform the sampling task on runtime based on the knowledge exchanged with other nodes and the spatial-temporal correlation among sensory data. Specially, we set a coverage table  $CovTable(v_k)$  stored at each node  $v_k$ , the size of which is dynamic. Each entry in the table is a two-tuple consisting of the grid cell identifier and the times (at least once) that the grid cell has been covered during some time period. Note that this table not only contains the grid cells covered by node  $v_k$ , but also contains the grid cells covered by other nodes, by exchanging coverage information with other nodes. At the beginning of each time period, we clear out the *coverage table* of each node. If some grid cell  $g_i$ , within which the node  $v_k$  is located, is not in its coverage table, or the coverage times of  $g_i$  is less than  $K$ , we make node  $v_k$  take a sample in  $g_i$ , and update its coverage table. Through cooperative sensing among mobile nodes, the sampling redundancy will be significantly reduced. The detailed algorithm is presented in Appendix A (see the supplementary file).

### 4.2 Cooperative Forwarding with Data Fusion

Before describing the cooperative forwarding schemes with data fusion, let us first illustrate the importance of coupling between opportunistic forwarding and data fusion in MONs.

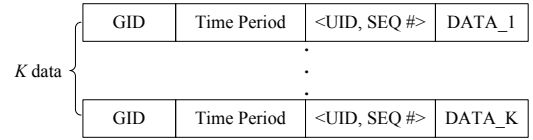
As depicted in Fig. 3, assume that two nodes  $u_1$  and  $u_2$  carry two correlated packets (two samples generated in the same grid cell within the same time period) at time  $t_1$ , and opportunistically forward the packets to a sink node  $S$ . Finally,  $S$  obtains the average of two samples. Fig. 3(a) illustrates the forwarding process without data fusion: when node  $u_1$  encounters node  $u_2$  at time  $t_2$ , they forward the data to each other; then at time  $t_3$ , node  $u_2$  encounters the sink node  $S$ , and delivers both the two packets  $A$  and  $B$  to  $S$ ;  $S$  take the average of two samples after it has received both the two original packets. Thus, the transmission overhead (i.e., the total number of transmissions) is 4. Next, let us see the forwarding process with data fusion as illustrated in Fig. 3(b): when  $u_1$  encounters  $u_2$ , they forward the data to each other, and store the fused result, namely a new packet  $C = (A + B)/2$ , instead of two original packets; then at time  $t_3$ ,  $u_2$  encounters  $S$ , and delivers the packet



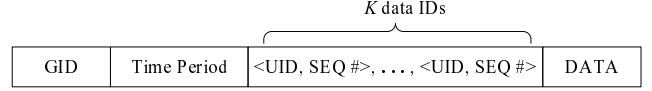
(a) Data forwarding without fusion. At time  $t_1$ , nodes  $u_1$  and  $u_2$  carry two correlated packets  $A$  and  $B$  respectively; at time  $t_2$ ,  $u_1$  forwards  $A$  to  $u_2$ , and  $u_2$  forwards  $B$  to  $u_1$ ; at time  $t_3$ ,  $u_2$  delivers  $A$  and  $B$  to the sink node  $S$ , and  $S$  obtain the average  $C$ .

(b) Data forwarding with fusion. At time  $t_1$ , nodes  $u_1$  and  $u_2$  carry two correlated packets  $A$  and  $B$  respectively; at time  $t_2$ ,  $u_1$  forwards  $A$  to  $u_2$ , and  $u_2$  forwards  $B$  to  $u_1$ , then both  $u_1$  and  $u_2$  obtain the fused packet  $C$ ; at time  $t_3$ ,  $u_2$  delivers  $C$  to the sink node  $S$ .

Fig. 3. Illustration of data forwarding with or without fusion.



(a)  $K$  original packets without data fusion.



(b) one packet fusing  $K$  data.

Fig. 4. Packet format with or without data fusion.

$C$  to  $S$ . Thus, the transmission overhead is 3, which is lower than that of forwarding process without data fusion.

Now we introduce the packet format of the two types of forwarding schemes. As depicted in Fig. 4(a), an original packet includes grid ID (GID), time period, generator ID (UID), locally unique sequence number (SEQ #) and the data. We consider such a node that has obtained  $K$  samples, either took by itself or received from other nodes, in the same grid cell within the same time period. If data fusion is not adopted in the forwarding process, this node needs to store  $K$  original packets as depicted in Fig. 4(a). In contrast, if data fusion is adopted, this node only needs to store one packet fusing  $K$  data (original packets), as depicted in Fig. 4(b). A more formal definition is given as follows:

**Definition 4** (Forwarding with Data Fusion). Let  $X = \{x_1, x_2, \dots, x_i\}$  and  $Y = \{y_1, y_2, \dots, y_j\}$  denote two packets having already fused  $i$  and  $j$  original packets respectively, and  $0 \leq i \leq j$ . When any node  $u_1$  with packet  $X$  encounters another node  $u_2$  with packet  $Y$ , both of them obtain a new fused packet by using a fusion function  $f(X, Y)$ . The

forwarding process follows the three rules below:

- (i) if  $X = Y$ , then nodes  $u_1$  and  $u_2$  do not forward any packet to each other;
- (ii) if  $X \subset Y$ , then  $u_2$  forwards a copy of packet  $Y$  to  $u_1$ ;
- (iii) if  $X \not\subseteq Y$ , then  $u_1$  forwards a copy of packet  $X$  to  $u_2$ , and  $u_2$  forwards a copy of packet  $Y$  to  $u_1$ .

In order to compare the relationship between  $X$  and  $Y$ , we exploit a space-efficient data structure for membership checking, i.e., a Bloom filter. The approach is similar to [35]. Each node uses a Bloom filter to represent its set of original packets, where each  $\langle \text{UID}, \text{SEQ} \# \rangle$  pair is used as input for the hash functions. By using this approach, only a small amount of data needs to be transmitted.

Many fusion functions, such as *averaging*, *summation*, *voting*, and *max/min*, can be applied in this scheme. According to the above rules, the transmission overhead in each forwarding process is at most 2. Thus the transmission overhead can be well controlled by data fusion. Note that some other fusion functions, such as *median* and *histogram*, cannot be applied in this scheme, because all original data should be reserved for computing the final result. In this paper, we focus only on the fusion functions that permit in-network processing.

## 5 COOPERATIVE FORWARDING SCHEMES

In this section, we present two cooperative forwarding schemes, and analyze their performance improvements in terms of delivery delay and transmission overhead.

### 5.1 Epidemic Routing with Fusion

Epidemic routing (ER) [13] has the minimum delivery delay under ideal conditions (unlimited resources such as bandwidth and buffer space), and thus can serve as an optimality baseline. Thus, we first integrate the ER scheme with data fusion, called ERF scheme. This integration is straightforward: whenever two nodes encounter, they perform forwarding process with fusion according to Definition 4. Since previous work on performance modeling of ER assumes that all packets are propagated individually, it is necessary to propose a new model to characterize the dissemination law of correlated packets, and analyze the effects of data fusion on the forwarding performance.

The propagation of one single packet can be modeled as spread of one infectious disease and the S-I-R model is often used for analyzing the propagation process [36]. The node is “infected” if it has already received a copy. The node is “susceptible” if it has not yet received a copy, but could potentially receive a copy from another node. We assume that all nodes move in the network independently and randomly, and the time duration between contacts of any two nodes follows an exponential distribution. The accuracy of this model has been shown for a number of random mobility models (e.g., random walk, random direction, and random waypoint) [31] and

some real mobility traces [37]. Let  $\beta$  denote the contact rate of the pairwise nodes. Let  $I(t)$  denote the number of infected nodes at time  $t$ , including the source node. Based on an ordinary differential equation (ODE) as a fluid limit of Markovian models, it is derived in [36] that:

$$I(t) = \frac{N}{1 + (N-1)e^{-\beta Nt}}.$$

Let  $F(t)$  denote the ratio of the nodes which carry a copy of the packet, which also equals to the probability that a given node has a copy of the packet. Then, we have

$$F(t) = \frac{I(t)}{N} = \frac{1}{1 + (N-1)e^{-\beta Nt}}.$$

Let  $I_i(t)$  denote the number of nodes which have  $i$  copies of  $K$  original packets under the ER scheme. It also equals to the number of nodes which have a packet fusing  $K$  original packets under the ERF scheme. Since each original packet is propagated individually, we can derive that

$$I_i(t) = \binom{K}{i} F^i(t) (1 - F(t))^{K-i} N. \quad (1)$$

**Lemma 1.** *Let's assume that there are  $K$  correlated packets originated from the mobile nodes in  $U$  at the same time. If the  $K$  packets are propagated individually using the ER scheme, then the total transmission overhead of  $K$  packets by time  $t$  is given by  $C(t) = KF(t)N - K$ .*

*Proof:* The detailed proof of this lemma is provided in Appendix B (see the supplementary file).  $\square$

**Theorem 1.** *Let's assume that there are  $K$  correlated packets originated from the mobile nodes in  $U$  at the same time. If the  $K$  packets are propagated using the ERF scheme, then the total transmission overhead by time  $t$ , denoted as  $C_f(t)$ , is given by the solution of the following system of rate equations:*

$$\begin{aligned} \frac{dC_f(t)}{dt} &= \sum_{0 \leq i < j \leq K} \left( 2 - \frac{\binom{j}{i}}{\binom{K}{i}} \right) I_i(t) I_j(t) \beta \\ &+ \sum_{0 < i = j \leq K} \left( 1 - \frac{1}{\binom{K}{i}} \right) I_i^2(t) \beta, \quad (2) \\ C_f(0) &= 0. \quad (3) \end{aligned}$$

*Proof:* The detailed proof of this theorem is provided in Appendix C (see the supplementary file).  $\square$

The above result, albeit quite useful in accurately predicting the transmission overhead of ERF, is not in closed form. This makes it difficult to theoretically compare the performances of ERF and ER. For this reason, we also derive a lower bound and an upper bound in closed form as follows.

**Theorem 2.** *The following lower bound and upper bound hold for the total transmission overhead of ERF:*

$$(1 - (1 - F(t))^K)N - K \leq C_f(t) \leq KF(t)N - K.$$

*Proof:* The detailed proof of this theorem is provided in Appendix D (see the supplementary file).  $\square$

The above theorem shows that the transmission overhead of ERF is less than or equal to that of ER. Now we analyze the delivery delay of the two schemes. Since we consider the scenario with  $K$  original packets propagating in the network, we define the delivery delay as the time within which all  $K$  original packets have been delivered to the monitoring center.

**Theorem 3.** *The expected delivery delay of ERF is equal to that of ER.*

*Proof:* The detailed proof of this theorem is provided in Appendix E (see the supplementary file).  $\square$

To get some insight into the above formulation, we will look in more detail into the special case of  $K = 2$ , as in Section 4.2. We present the theoretical analysis and simulation results under the Random Way Point mobility model to compare the transmission overhead and delivery delay of the BSW and BSWF schemes in Appendix F (see the supplementary file).

## 5.2 Spray-and-Wait with Fusion

Spray-and-wait [19] is considered to be one of the most efficient methods to reduce the large overhead of ER without incurring significant delay penalties. The rationale is: the source node distributes only a small number of copies to different relay nodes using epidemic-based schemes (spray phase), and then each relay node performs “direct transmission”, i.e., only forwards the packet to its destination (wait phase). Among various spraying heuristics, *Binary Spraying* is proved to be optimal [19], when node movement is independent and identically distributed. We give a formal definition as follows:

**Definition 5** (Binary Spray-and-Wait (BSW)). *When a new packet is generated at a source node, this node also creates  $L > 1$  “forwarding tokens” for this packet. A forwarding token implies that the node that owns it can spawn and forward an additional copy of the given packet. The forwarding process consists of the following two phases:*

- *binary spray phase: if a node (either the source or a relay), carrying a packet copy and  $l > 1$  forwarding tokens for this packet, encounters a node with no copy of the packet, it spawns and forwards a copy of that packet to the second node; it also hands over  $\lfloor l/2 \rfloor$  forwarding tokens to that node and keeps the rest  $\lfloor l/2 \rfloor$  for itself;*
- *wait phase: when a node has a packet copy and  $l = 1$  forwarding tokens for this packet, then it can only forward this packet to the destination itself.*

If we consider integrating BSW with data fusion, an important question must be answered: how many forwarding tokens should be assigned to the nodes for the new fused packet? Compared to the epidemic-based forwarding schemes, BSW can reduce the transmission overhead at the cost of increasing delivery delay. Thus,

we prefer to design such a BSW scheme with data fusion that the delivery delay can be reduced while not increasing the transmission overhead. According to this criterion, we assign proper number of forwarding tokens to the nodes. The detailed rules are defined as follows.

**Definition 6** (Binary Spray-and-Wait with Fusion (BSWF)). *Let  $X = \{x_1, x_2, \dots, x_i\}$  and  $Y = \{y_1, y_2, \dots, y_j\}$  denote two packets having already fused  $i$  and  $j$  original packets respectively, and  $0 \leq i \leq j$ . Assume that one node  $u_1$  carries a copy of packet  $X$  and  $l_1 \geq 0$  forwarding tokens for  $X$ , and another node  $u_2$  carries a copy of packet  $Y$  and  $l_2 \geq 0$  forwarding tokens for  $Y^*$ . When  $u_1$  encounters  $u_2$ , they forward data following the rules in Table 1, based on the relationship between  $X$  and  $Y$ , and the number of copies.*

In order to facilitate comparing the performances of BSW and BSWF in terms of delivery delay and transmission overhead, we will use the variant of Definition 5 to characterize BSW without data fusion by considering the scenario with  $K$  original packets propagating in the network individually.

**Definition 7** (BSW without Fusion (a variant of Definition 5)). *Let  $X = \{x_1, x_2, \dots, x_i\}$  and  $Y = \{y_1, y_2, \dots, y_j\}$  denote two sets of original packets, and  $0 \leq i \leq j$ . Assume that one node  $u_1$  carries a set of packets  $X$  and  $l_1 \geq 0$  forwarding tokens for each original packet in  $X$ , and another node  $u_2$  carries a set of packets  $Y$  and  $l_2 \geq 0$  forwarding tokens for each original packet in  $Y$ . When  $u_1$  encounters  $u_2$ , they forward data following the rules in Table 2, based on the relationship between  $X$  and  $Y$ , and the number of copies.*

More detailed comparison between BSW and BSWF is presented in Appendix G (see the supplementary file).

**Theorem 4.** *The transmission overhead of BSWF is lower than or equal to that of BSW.*

*Proof:* The detailed proof of this theorem is provided in Appendix H (see the supplementary file).  $\square$

**Theorem 5.** *The expected delivery delay of BSWF is lower than that of BSW.*

*Proof:* The detailed proof of this theorem is provided in Appendix I (see the supplementary file).  $\square$

We perform simulations under the Random Way Point mobility model to compare the transmission overhead and delivery delay of the BSW and BSWF schemes. The detailed results are presented in Appendix J (see the supplementary file).

## 6 PERFORMANCE EVALUATION

### 6.1 Evaluation Environment and Settings

We evaluate the performance of COUPON based on the real mobility traces collected from KAIST [24]. Altogether 92 daily trajectories are collected by participants with

\*. Only when  $i = 0$  or  $j = 0$ , we have  $l_1 = 0$  or  $l_2 = 0$ ; otherwise we have  $l_1 \geq 1$  and  $l_2 \geq 1$ .

TABLE 1  
Forwarding rules of BSWF.

Rules	Conditions	Operations
(i)	$X \subset Y$ $l_2 > 1$	$u_2$ spawns and forwards a copy of $Y$ to $u_1$ ; $u_1$ and $u_2$ hold $\lfloor (l_1 + l_2)/2 \rfloor$ and $\lceil (l_1 + l_2)/2 \rceil$ forwarding tokens for a new fused packet $Z = f(X, Y) = Y$ , respectively.
(ii-i)	$X \not\subseteq Y, X \cap Y \neq \emptyset,$ $l_1 > 1, l_2 > 1$	$u_1$ spawns and forwards a copy of $X$ to $u_2$ , $u_2$ spawns and forwards a copy of $Y$ to $u_1$ ; $u_1$ and $u_2$ hold $\lfloor (l_1 + l_2)/2 \rfloor$ and $\lceil (l_1 + l_2)/2 \rceil$ forwarding tokens for a new fused packet $Z = f(X, Y)$ , respectively.
(ii-ii)	$X \not\subseteq Y, X \cap Y \neq \emptyset,$ $l_1 = 1, l_2 > 1$	$u_2$ spawns and forwards a copy of $Y$ to $u_1$ ; $u_1$ holds $\lfloor (1 + l_2)/2 \rfloor$ forwarding tokens for a new fused packet $Z = f(X, Y)$ , and $u_2$ holds $\lceil (1 + l_2)/2 \rceil$ forwarding tokens for the packet $Y$ .
(ii-iii)	$X \not\subseteq Y, X \cap Y \neq \emptyset,$ $l_1 > 1, l_2 = 1$	$u_1$ spawns and forwards a copy of $X$ to $u_2$ ; $u_2$ holds $\lfloor (l_1 + 1)/2 \rfloor$ forwarding tokens for a new fused packet $Z = f(X, Y)$ , and $u_1$ holds $\lceil (l_1 + 1)/2 \rceil$ forwarding tokens for the packet $X$ .
(iii-i)	$X \not\subseteq Y, X \cap Y = \emptyset$ $l_1 > 1, l_2 > 1$	$u_1$ spawns and forwards a copy of $X$ to $u_2$ , $u_2$ spawns and forwards a copy of $Y$ to $u_1$ ; $u_1$ and $u_2$ hold $\lfloor \max(l_1, l_2)/2 \rfloor$ and $\lceil \max(l_1, l_2)/2 \rceil$ forwarding tokens for a new fused packet $Z = f(X, Y)$ , respectively.
(iii-ii)	$X \not\subseteq Y, X \cap Y = \emptyset$ $l_1 = 1, l_2 > 1$	$u_2$ spawns and forwards a copy of $Y$ to $u_1$ ; $u_1$ holds $\lfloor l_2/2 \rfloor$ forwarding tokens for a new fused packet $Z = f(X, Y)$ , and $u_2$ holds $\lceil l_2/2 \rceil$ forwarding tokens for the packet $Y$ .
(iii-iii)	$X \not\subseteq Y, X \cap Y = \emptyset$ $l_1 > 1, l_2 = 1$	$u_1$ spawns and forwards a copy of $X$ to $u_2$ ; $u_2$ holds $\lfloor l_1/2 \rfloor$ forwarding tokens for a new fused packet $Z = f(X, Y)$ , and $u_1$ holds $\lceil l_1/2 \rceil$ forwarding tokens for the packet $X$ .
(iv)	others not applicable	$u_1$ and $u_2$ do not forward any packet or hand over any forwarding token to each other.

TABLE 2  
Forwarding rules of BSW.

Rules	Conditions	Operations
(i)	$X \subset Y$ $l_2 > 1$	$u_2$ spawns and forwards a copy of each packet in $Y - X$ to $u_1$ ; $u_1$ hands over $\lfloor l_2/2 \rfloor$ forwarding tokens for each packet in $Y - X$ to $u_2$ , and keeps the rest $\lceil l_2/2 \rceil$ .
(ii-i)	$X \not\subseteq Y, X \cap Y \neq \emptyset,$ $l_1 > 1, l_2 > 1$	$u_1$ spawns and forwards a copy of each packet in $X - X \cap Y$ to $u_2$ , $u_2$ spawns and forwards a copy of each packet in $Y - X \cap Y$ to $u_1$ ; $u_1$ hands over $\lfloor l_1/2 \rfloor$ tokens for each packet in $X - X \cap Y$ to $u_2$ , and keeps the rest $\lceil l_1/2 \rceil$ ; $u_2$ hands over $\lfloor l_2/2 \rfloor$ tokens for each packet in $Y - X \cap Y$ to $u_1$ , and keeps the rest $\lceil l_2/2 \rceil$ .
(ii-ii)	$X \not\subseteq Y, X \cap Y \neq \emptyset,$ $l_1 = 1, l_2 > 1$	$u_2$ spawns and forwards a copy of each packet in $Y - X \cap Y$ to $u_1$ ; $u_2$ hands over $\lfloor l_2/2 \rfloor$ forwarding tokens for each packet in $Y - X \cap Y$ to $u_1$ , and keeps the rest $\lceil l_2/2 \rceil$ .
(ii-iii)	$X \not\subseteq Y, X \cap Y \neq \emptyset,$ $l_1 > 1, l_2 = 1$	$u_1$ spawns and forwards a copy of each packet in $X - X \cap Y$ to $u_2$ ; $u_1$ hands over $\lfloor l_1/2 \rfloor$ forwarding tokens for each packet in $X - X \cap Y$ to $u_2$ , and keeps the rest $\lceil l_1/2 \rceil$ .
(iii-i)	$X \not\subseteq Y, X \cap Y = \emptyset,$ $l_1 > 1, l_2 > 1$	$u_1$ spawns and forwards a copy of each packet in $X$ to $u_2$ , and $u_2$ spawns and forwards a copy of each packet in $Y$ to $u_1$ ; $u_1$ hands over $\lfloor l_1/2 \rfloor$ tokens for each packet in $X$ to $u_2$ , and keeps the rest $\lceil l_1/2 \rceil$ ; $u_2$ hands over $\lfloor l_2/2 \rfloor$ tokens for each packet in $Y$ to $u_1$ , and keeps the rest $\lceil l_2/2 \rceil$ .
(iii-ii)	$X \not\subseteq Y, X \cap Y = \emptyset,$ $l_1 = 1, l_2 > 1$	$u_2$ spawns and forwards a copy of each packet in $Y$ to $u_1$ ; $u_2$ hands over $\lfloor l_2/2 \rfloor$ forwarding tokens for each packet in $Y$ to $u_1$ , and keeps the rest $\lceil l_2/2 \rceil$ .
(iii-iii)	$X \not\subseteq Y, X \cap Y = \emptyset,$ $l_1 > 1, l_2 = 1$	$u_1$ spawns and forwards a copy of each packet in $X$ to $u_2$ ; $u_1$ hands over $\lfloor l_1/2 \rfloor$ forwarding tokens for each packet in $X$ to $u_2$ , and keeps the rest $\lceil l_1/2 \rceil$ .
(iv)	others not applicable	$u_1$ and $u_2$ do not forward any packet or hand over any forwarding token to each other.

GPS hand-held receivers at every 10 seconds. To account for GPS errors, a position is recomputed at every 30 seconds by averaging three samples over that 30 second period. The minimum time duration of the 92 trajectories is 15150 seconds length. In this work, we exploit the 92 trajectories during the first 14400 seconds (i.e., 4 hours) for evaluation. All trajectories are mapped into a two dimensional area. We divide the sensing field into  $200 \times 200\text{m}^2$  grid cells. This size is the same as the setting in [5] for  $CO_2$  concentration monitoring according to actual measurements of the spatio-temporal sensitivity. One sink node is placed at the center of the sensing field. The time period and sampling period are set to be  $T_p = 1800\text{s}$  and  $T_s = 30\text{s}$ . Five different coverage constraints ( $K = 1 \sim 5$ ) are set for each effectively covered grid cell. The communication range of each node is set to be 70m, which is the same as the setting in [11],

[16], [17]. The TTL constraint for each grid cell is set to be 3600s.

## 6.2 Evaluation Results of Cooperative Sensing

Table 3 shows the number of effectively covered grid cells with various values of  $K$  within different time periods. The larger the value of  $K$ , the more coverage times are constrained for each effectively covered grid cell. Thus, we can see that the number of effectively covered grid cells decreases with the increase of  $K$ . Fig. 5 compares the total number of samplings performed in all effectively covered grid cells within 8 time periods by using non-cooperative sensing and cooperative sensing. We can see that the cooperative sensing scheme can significantly reduce the number of samplings, thus reduce the energy consumption in the sensing process. Specially,



TABLE 3

The number of effectively covered grid cells with various values of  $K$  within different time Periods (“P.”, for short).

$K$	# of effectively covered grids							
	1st P.	2nd P.	3rd P.	4th P.	5th P.	6th P.	7th P.	8th P.
1	121	124	141	128	96	77	132	110
2	80	74	97	81	62	63	88	76
3	64	65	84	63	59	57	76	67
4	54	56	71	53	51	53	69	62
5	44	50	53	49	46	49	64	60

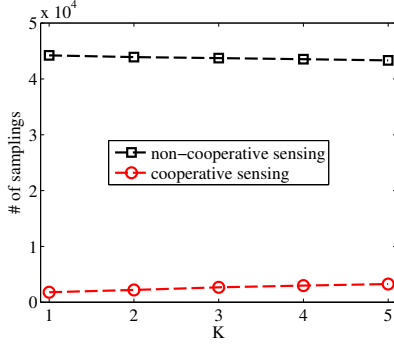


Fig. 5. The total number of samplings performed in all effectively covered grid cells within 8 time periods, with various values of  $K$ .

the total number of samplings of the cooperative sensing scheme at  $K = 5$  reduces by 93% compared with the non-cooperative scheme.

### 6.3 Evaluation Results of Cooperative Forwarding

We compare and evaluate the performances of four forwarding schemes in terms of three metrics as follows.

**Average number of effectively delivered grid cells:** It denotes how many effectively covered grid cells can be effectively delivered within one time period on average. Fig. 6(a) compares the results with various values of  $K$  by using four forwarding schemes. The number of effectively delivered grid cells decreases with the increase of  $K$  by using any forwarding scheme. As expected, both ER and ERF have the largest number of effectively delivered grid cells. In general, more grid cells can be effectively delivered by using BSWF compared with BSW. Specially, the delivery ratio of BSWF at  $K = 5$  increases by 16% compared with BSW.

**Average delivery delay:** Fig. 6(b) compares the average delivery delay of each effectively delivered grid cell by using four forwarding schemes. As expected, both ER and ERF have the lowest delivery delay. In general, by using data fusion, BSWF can achieve lower delivery delay than BSW. Specially, the average delivery delay of BSWF at  $K = 5$  reduces by 5% compared with BSW.

**Average transmission overhead:** Fig. 6(c) compares the average transmission overhead of each effectively delivered grid cell by using four forwarding schemes. As expected, ER has the highest transmission overhead. In general, by using data fusion, ERF and BSWF can achieve lower transmission overhead than ER and BSW

respectively. Specially, the average transmission overheads of ERF and BSWF at  $K = 5$  reduce by 78% and 32% compared with ER and BSW respectively. In addition, the transmission overhead of ER increases linearly with the value of  $K$ . By contrast, the transmission overheads of the other three schemes increase slowly with the value of  $K$ .

In summary, the above simulation results verify that: 1) the ERF scheme can significantly reduce the transmission overhead of the ER scheme with the same delivery ratio and delivery delay; 2) the BSWF scheme can increase the delivery ratio of the BSW scheme, and reduce both the delivery delay and transmission overhead; 3) the ERF and BSWF schemes can achieve excellent tradeoff among delivery ratio, delivery delay and transmission overhead by integrating the forwarding schemes with data fusion.

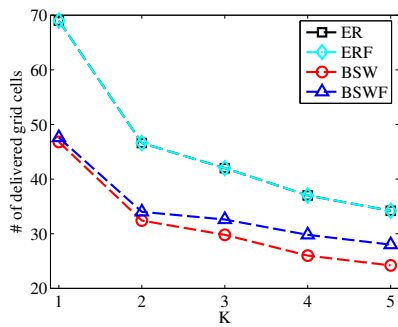
## 7 CONCLUSIONS AND FUTURE WORK

In this paper, we proposed COUPON, a novel cooperative sensing and data forwarding framework for building sensing maps while satisfying specific sensing quality with low delay and energy consumption. The cooperative sensing scheme can eliminate sampling redundancy and hence save energy. Two cooperative data forwarding schemes, ERF and BSWF, can achieve better tradeoff between delivery delay and transmission overhead by taking advantage of data fusion. We evaluated our schemes using real mobile traces. Extensive simulations demonstrated that the cooperative sensing scheme could reduce the number of samplings by 93% compared with the non-cooperative scheme; ERF could reduce the transmission overhead by 78% compared with ER; BSWF could increase the delivery ratio by 16%, and reduce the delivery delay and transmission overhead by 5% and 32% respectively, compared with BSW.

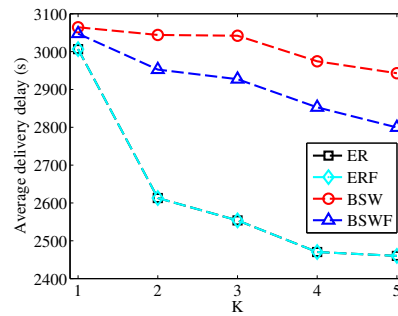
Although we only focus on two opportunistic forwarding protocols in this paper, we believe that our derived dissemination law of correlated packets can serve as fundamental guidelines on integrating other opportunistic forwarding protocols with data fusion. In the future, we are interested in combining some other opportunistic forwarding protocols, such as context-aware and social-based forwarding, with in-network processing for further improving network performance in different application scenarios. We will also explore the possibility to combine opportunistic forwarding protocols with more complex in-network processing.

## ACKNOWLEDGMENT

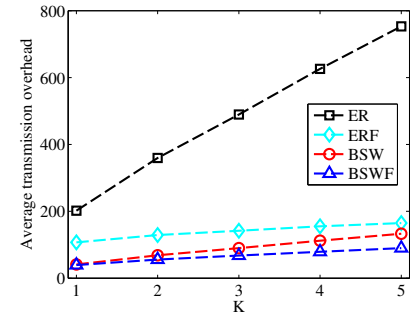
This work is supported by the National Natural Science Foundation of China under Grant No.61332005 and No.61272517, the National Science Fund for Distinguished Young Scholars under Grant No.60925010, the Funds for Creative Research Groups of China under Grant No.61121001, the Specialized Research Fund for the Doctoral Program of Higher Education under Grant No.20120005130002.



(a) The average number of effectively delivered grid cells within one time period.



(b) The average delivery delay of each effectively delivered grid cell.



(c) The average transmission overhead of each effectively delivered grid cell.

Fig. 6. Performance comparisons of four forwarding schemes.

## REFERENCES

- [1] P. Dutta, P. Aoki, N. Kumar, A. Mainwaring, C. Myers, W. Willett, and A. Woodruff, "Common sense: participatory urban sensing using a network of handheld air quality monitors," in *Proc. of ACM SenSys*, 2009, pp. 349–350.
- [2] M. Stevens and E. D'Hondt, "Crowdsourcing of pollution data using smartphones," in *Workshop on Ubiquitous Crowdsourcing*, 2010.
- [3] R. Rana, C. Chou, S. Kanhere, N. Bulusu, and W. Hu, "Ear-phone: an end-to-end participatory urban noise mapping system," in *Proc. of ACM/IEEE IPSN*, 2010, pp. 105–116.
- [4] X. Yu, H. Zhao, L. Zhang, S. Wu, B. Krishnamachari, and V. Li, "Cooperative sensing and compression in vehicular sensor networks for urban monitoring," in *Proc. of IEEE ICC*, 2010, pp. 1–5.
- [5] S. Hu, Y. Wang, C. Huang, and Y. Tseng, "A vehicular wireless sensor network for CO<sub>2</sub> monitoring," *IEEE Sensors*, pp. 1498–1501, 2009.
- [6] A. Campbell, S. Eisenman, N. Lane, E. Miluzzo, and R. Peterson, "People-centric urban sensing," in *Second ACM/IEEE International Conference on Wireless Internet (WiCon)*, 2006, pp. 18–31.
- [7] M. Demirbas, M. A. Bayir, C. G. Akcora, Y. S. Yilmaz, and H. Ferhatosmanoglu, "Crowd-sourced sensing and collaboration using twitter," in *Proc. of IEEE WoWMoM*, 2010, pp. 1–9.
- [8] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: Current state and future challenges," *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 32–39, 2011.
- [9] A. Campbell, S. Eisenman, N. Lane, E. Miluzzo, R. Peterson, H. Lu, X. Zheng, M. Musolesi, K. Fodor, and G. Ahn, "The rise of people-centric sensing," *IEEE Internet Comp.*, vol. 12, no. 4, pp. 12–21, 2008.
- [10] M. Srivastava, T. Abdelzaher, and B. K. Szymanski, "Human-centric sensing," *Philosophical Trans. of the Royal Society, special issue on Wireless Sensor Networks*, 2011.
- [11] E. Ngai, M. B. Srivastava, and J. Liu, "Context-aware sensor data dissemination for mobile users in remote areas," in *Proc. of IEEE INFOCOM*, 2012, pp. 2711–2715.
- [12] M. Y. S. Uddin, H. Wang, F. Saremi, G.-J. Qi, T. Abdelzaher, and T. Huang, "Photonet: a similarity-aware picture delivery service for situation awareness," in *Proc. of IEEE RTSS*, 2011, pp. 317–326.
- [13] A. Vahdat, D. Becker *et al.*, "Epidemic routing for partially connected ad hoc networks," Technical Report CS-200006, Tech. Rep., 2000.
- [14] L. Pelusi, A. Passarella, and M. Conti, "Opportunistic networking: data forwarding in disconnected mobile ad hoc networks," *IEEE Commun. Mag.*, vol. 44, no. 11, pp. 134–141, 2006.
- [15] Z. Zhang, "Routing in intermittently connected mobile ad hoc networks and delay tolerant networks: overview and challenges," *IEEE Commun. Surveys*, vol. 8, no. 1, pp. 24–37, 2006.
- [16] E.-H. Ngai, H. Huang, J. Liu, and M. B. Srivastava, "Oppsense: Information sharing for mobile phones in sensing field with data repositories," in *Proc. of IEEE SECON*, 2011, pp. 107–115.
- [17] U. Park and J. Heidemann, "Data muling with mobile phones for sensornets," in *Proc. of ACM SenSys*, 2011, pp. 162–175.
- [18] H. Luo, Y. Liu, and S. Das, "Routing correlated data in wireless sensor networks: A survey," *IEEE Network*, vol. 21, no. 6, pp. 40–47, 2007.
- [19] T. Spyropoulos, K. Psounis, and C. Raghavendra, "Efficient routing in intermittently connected mobile networks: the multiple-copy case," *IEEE/ACM Trans. on Networking*, vol. 16, no. 1, pp. 77–90, 2008.
- [20] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing," in *Proc. of ACM MobiCom*, 2012.
- [21] D. Zhao, X.-Y. Li, and H. Ma, "How to crowdsource tasks truthfully without sacrificing utility: online incentive mechanisms with budget constraint," in *Proc. of IEEE INFOCOM*, 2014.
- [22] J. Shi, R. Zhang, Y. Liu, and Y. Zhang, "Prisense: privacy-preserving data aggregation in people-centric urban sensing systems," in *Proc. of IEEE INFOCOM*, 2010, pp. 1–9.
- [23] Q. Li and G. Cao, "Efficient and privacy-preserving data aggregation in mobile sensing," in *Proc. of IEEE ICNP*, 2012, pp. 1–10.
- [24] K. Lee, S. Hong, S. Kim, I. Rhee, and S. Chong, "Slaw: A mobility model for human walks," in *Proc. of IEEE INFOCOM*, 2009, pp. 855–863.
- [25] D. Zhao, H. Ma, and S. Tang, "Coupon: cooperatively building sensing maps in mobile opportunistic networks," in *Proc. of IEEE MASS*, 2013, pp. 295–303.
- [26] D. Zhao, H. Ma, L. Liu, and J. Zhao, "On opportunistic coverage for urban sensing," in *Proc. of IEEE MASS*, 2013, pp. 231–239.
- [27] "Sensorly," <http://www.sensorly.com>.
- [28] B. Hull, V. Bychkovsky, Y. Zhang, K. Chen, M. Goraczko, A. Miu, E. Shih, H. Balakrishnan, and S. Madden, "Cartel: a distributed mobile sensor computing system," in *Proc. of ACM SenSys*, 2006, pp. 125–138.
- [29] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, "The pothole patrol: using a mobile sensor network for road surface monitoring," in *Proc. of ACM MobiSys*, 2008.
- [30] S. Mathur, T. Jin, N. Kasturirangan, J. Chandrasekaran, W. Xue, M. Gruteser, and W. Trappe, "Parknet: drive-by sensing of roadside parking statistics," in *Proc. of ACM MobiSys*, 2010, pp. 123–136.
- [31] R. Groenevelt, P. Nain, and G. Koole, "The message delay in mobile ad hoc networks," *Performance Evaluation*, vol. 62, no. 1-4, pp. 210–228, 2005.
- [32] A. Lindgren, A. Doria, and O. Schelén, "Probabilistic routing in intermittently connected networks," *ACM Mobile Comp. and Commun. Rev.*, vol. 7, no. 3, pp. 19–20, 2003.
- [33] S. Kumar, T. Lai, and J. Balogh, "On k-coverage in a mostly sleeping sensor network," in *Proc. of ACM MobiCom*, 2004, pp. 144–158.
- [34] M. Hefeeda and M. Bagheri, "Randomized k-coverage algorithms for dense sensor networks," in *Proc. of IEEE INFOCOM*, 2007, pp. 2376–2380.
- [35] U. Lee, E. Magistretti, M. Gerla, P. Bellavista, and A. Corradi, "Dissemination and harvesting of urban data using vehicular sensing platforms," *IEEE Trans. on Vehicular Technology*, vol. 58, no. 2, pp. 882–901, 2009.
- [36] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance modeling of epidemic routing," *Comp. Net.*, vol. 51, no. 10, pp. 2867–2891, 2007.
- [37] N. Banerjee, M. Corner, D. Towsley, and B. Levine, "Relays, base stations, and meshes: enhancing mobile networks with infrastructure," in *Proc. of ACM MobiCom*, 2008, pp. 81–91.

## APPENDIX A

### COOPERATIVE SENSING ALGORITHM

**Algorithm 1** Cooperative Sensing Algorithm (for each node  $v_k$ )

---

```

1: for  $t = T_s$  to  $T$  do
2:   if  $t \% T_p = 0$  then
3:     Set  $CovTable(v_k) = \text{NULL}$ ;
4:   end if
5:   Obtain the position of  $v_k$ :  $L_k(t)$ ;
6:   Obtain the grid identifier  $g_i$ , which is covered by
    $v_k$ , i.e.,  $C_i(L_k(t)) = 1$ ;
7:   if  $g_i$  is not in  $CovTable(v_k)$  then
8:     Take a sample in  $g_i$ ;
9:     Add  $g_i$  to  $CovTable(v_k)$ ;
10:    Set the coverage times of  $g_i$  to 1;
11:   end if
12:   if  $g_i$  is in  $CovTable(v_k)$ , and the coverage times of
    $g_i$  is less than  $K$  then
13:     Take a sample in  $g_i$ ;
14:     Add 1 to the coverage times of  $g_i$ ;
15:   end if
16: end for

```

---

We outline the *cooperative sensing* algorithm in Algorithm 1. Take node  $v_k$  as an example. At each time slot, it performs cooperative sensing algorithm. If the current time slot is just at the beginning of one time period, node  $v_k$  clears out its *coverage table* (line 2-4). Then we do the following steps: if the current grid cell  $g_i$ , within which the node  $v_k$  is located, is not in its *coverage table*, or the coverage times of  $g_i$  is less than  $K$ , we make node  $v_k$  take a sample in  $g_i$ , and update its *coverage table* (line 6-18). Through cooperative sensing among mobile nodes, the sampling redundancy will be significantly reduced.

## APPENDIX B

### PROOF OF LEMMA 1

Each packet will have  $F(t)N$  copies in the network by time  $t$ . Without data fusion, each copy of packets means one transmission except for the  $K$  original packets. Thus, we have that:  $C(t) = KF(t)N - K$ .

## APPENDIX C

### PROOF OF THEOREM 1

A unified framework based on ODEs has been used to study ER and its variations [36]. Here, we extend this framework to analyze the transmission overhead of ERF.  $\frac{dC_f(t)}{dt}$  means the increasing rate of the total transmission overhead, which is caused by two cases: 1) a class of nodes with a packet fusing  $i$  original packets encounter another class of nodes with a packet fusing  $j$  original packets,  $i \neq j$ ; 2) two classes of nodes with a packet fusing the same number of original packets but with

different data encounter with each other. We discuss the two cases as follows.

**Case 1:** In this case, the increasing rate of the transmission overhead can be expressed by:

$$\frac{dC_{f1}(t)}{dt} = \sum_{0 \leq i < j \leq K} I_i(t)I_j(t)\mathbf{E}[Q_{ij}]\beta,$$

where  $Q_{ij}$  denotes the transmission overhead when a node with a packet fusing  $i$  original packets encounters another node with a packet fusing  $j$  original packets, and  $\mathbf{E}[Q_{ij}]$  denotes its expectation.

Let  $X = \{x_1, x_2, \dots, x_i\}$  and  $Y = \{y_1, y_2, \dots, y_j\}$  denote two packets fusing  $i$  and  $j$  original packets respectively, and  $0 \leq i < j \leq K$ . When two nodes with the two packets encounter, the transmission overhead is given by:

$$Q_{ij} = \begin{cases} 1, & \text{if } X \subset Y; \\ 2, & \text{otherwise.} \end{cases}$$

It can be derived that:

$$P(X \subset Y) = \binom{j}{i} / \binom{K}{i}.$$

$$\begin{aligned} \mathbf{E}[Q_{ij}] &= 1 \times P(X \subset Y) + 2 \times (1 - P(X \subset Y)) \\ &= 2 - \binom{j}{i} / \binom{K}{i}. \end{aligned}$$

**Case 2:** In this case, we divide the nodes with a packet fusing  $i$  original packets into  $\binom{K}{i}$  classes, where the nodes from the same class have a packet fusing the same data, and the nodes from different classes have a packet fusing different data. Each class contains  $I_i(t) / \binom{K}{i}$  nodes. When two nodes from the same class encounter, since they have the same data, no transmission is needed; when two nodes from different classes encounter, the transmission overhead is 2. There are  $\binom{K}{2}$  cases that two nodes from different classes encounter. Thus, the increasing rate of the transmission overhead can be expressed by:

$$\begin{aligned} \frac{dC_{f2}(t)}{dt} &= \sum_{0 < i = j \leq K} \binom{K}{2} \left( \frac{I_i(t)}{\binom{K}{i}} \right)^2 \times 2\beta \\ &= \sum_{0 < i = j \leq K} \left( 1 - \frac{1}{\binom{K}{i}} \right) I_i^2(t)\beta. \end{aligned}$$

Combining the above two cases together, we can obtain the increasing rate of the total transmission overhead:

$$\frac{dC_f(t)}{dt} = \frac{dC_{f1}(t)}{dt} + \frac{dC_{f2}(t)}{dt}.$$

Thus, Eq. (2) can be derived. The initial condition (3) simply states that the transmission overhead is 0 at first.

## APPENDIX D

### PROOF OF THEOREM 2

Assume that one node has a packet fusing  $i$  original packets  $X = \{x_1, x_2, \dots, x_i\}$ . We consider two extreme cases: 1) this node receives the packet  $X$  directly from another node, so it only needs one transmission; 2) this node receives only one single original packet each time from other nodes, so it needs to obtain the packet  $X$  through  $i$  transmissions altogether. Thus, we can derive that:

$$\sum_{i=1}^K I_i(t) - K \leq C_f(t) \leq \sum_{i=1}^K iI_i(t) - K. \quad (4)$$

According to the binomial theorem and Eq. (1), it can be derived that:

$$\sum_{i=1}^K I_i(t) = \sum_{i=0}^K I_i(t) - I_0(t) = N - I_0(t) = (1 - (1 - F(t))^K)N. \quad (5)$$

Combing the equations (1),(4) and (5), we can easily obtain the lower bound and upper bound of  $C_f(t)$ .

## APPENDIX E

### PROOF OF THEOREM 3

From Definition 4, we see that during each forwarding process, the data fusion cannot increase or decrease the propagation speed of each original packet, thus it has not any effect on the expected delivery delay.

## APPENDIX F

### SPECIAL CASE ANALYSIS ( $K=2$ )

Let's consider the example in Fig. 3. We assume that there are two correlated packet  $A$  and  $B$ , which denote two samples taken by two different nodes in the same grid cell within the same time period. When a node with a copy of packet  $A$  encounters another node with a copy of packet  $B$ , both the two nodes fuse the two packets into a new packet  $C$  (e.g., take an average). Let  $I_A(t)$  and  $I_B(t)$  denote the number of nodes that only carry a copy of  $A$  or  $B$ . Let  $I_C(t)$  denote the number of nodes that carry a fused packet  $C$ , and  $I_0(t)$  denote the number of nodes that do not carry any one copy of  $A$ ,  $B$  or  $C$ . Then, we have

$$I_A(t) = I_B(t) = \frac{1}{2}I_1(t) = F(t)(1 - F(t))N,$$

$$I_C(t) = I_2(t) = F^2(t)N,$$

$$I_0(t) = (1 - F(t))^2N.$$

According to Theorem 1, we can obtain that:

$$\frac{dC_f(t)}{dt} = F(t)(1 - F(t))(F^2(t) - F(t) + 2)N^2\beta. \quad (6)$$

Now, we use Markov chain to derive the solution of  $C(t)$  and  $C_f(t)$ , and their relationship. This approach is more straightforward. For the ER scheme, we classify

TABLE 4

Transition rules among various states under the ER scheme.

$\wedge$	0	A	B	A&B
0	0 (0)	A (1)	B (1)	A&B (2)
A	A (1)	A (0)	A&B (2)	A&B (1)
B	B (1)	A&B (2)	B (0)	A&B (1)
A&B	A&B (2)	A&B (1)	A&B (1)	A&B (0)

TABLE 5

Transition rules among various states under the ERF scheme.

$\wedge$	0	A	B	C
0	0 (0)	A (1)	B (1)	C (1)
A	A (1)	A (0)	C (2)	C (1)
B	B (1)	C (2)	B (0)	C (1)
C	C (2)	C (1)	C (1)	C (0)

all nodes into four states:  $A$ ,  $B$ ,  $A&B$ , and  $0$ . The node with only a copy of  $A$  or  $B$  is in the state  $A$  or  $B$ ; the node with copies of  $A$  and  $B$  simultaneously is in the state  $A&B$ , and the node without any copy of  $A$  or  $B$  is the state  $0$ . Similarly, for the ERF scheme, all nodes are classified into four states:  $A$ ,  $B$ ,  $C$ , and  $0$ . Let symbol " $\wedge$ " denote the transition rules among various states when two nodes encounter with each other. All transition rules under the ER and ERF schemes are listed in Table 4 and Table 5, respectively. The number in the bracket denotes the corresponding transmission overhead whenever the states change.

According to Table 4 and Table 5, we can obtain that:

$$\begin{aligned} \frac{dC(t)}{dt} &= I_0(t)I_A(t)\beta + I_0(t)I_B(t)\beta + 2I_0(t)I_{AB}(t)\beta \\ &\quad + 2I_A(t)I_B(t)\beta + I_A(t)I_{AB}(t)\beta + I_B(t)I_{AB}(t)\beta \\ &= 2F(t)(1 - F(t))N^2\beta. \end{aligned} \quad (7)$$

$$\begin{aligned} \frac{dC_f(t)}{dt} &= I_0(t)I_A(t)\beta + I_0(t)I_B(t)\beta + I_0(t)I_C(t)\beta \\ &\quad + 2I_A(t)I_B(t)\beta + I_A(t)I_C(t)\beta + I_B(t)I_C(t)\beta \\ &= F(t)(1 - F(t))(F^2(t) - F(t) + 2)N^2\beta. \end{aligned} \quad (8)$$

We see that the above result of  $\frac{dC_f(t)}{dt}$  is consistent with Eq. (6), which is derived by Theorem 1.

According to the above two equations, we can derive the relationship between  $C(t)$  and  $C_f(t)$ :

$$C_f(t) = C(t) - \int F^2(t)(1 - F(t))^2N^2\beta dt.$$

In Fig. 7 we compare simulation results for the transmission overhead of ER and ERF schemes against analytical values calculated according to Eq. (7) and Eq. (8), as a function of time. We observe that theoretical and simulation results are closely matched, which confirm the accuracy of our analytical model.

TABLE 6  
Comparison between BSW and BSWF.

Rules	Schemes	Packets (# of forwarding tokens)		Current overhead	Potential overhead
		$u_1$	$u_2$		
(i)	BSW	$X(l_1), Y - X(\lfloor l_2/2 \rfloor)$	$X(l_2), Y - X(\lceil l_2/2 \rceil)$	$\ Y - X\ $	$(l_1 + l_2 - 2)\ X\  + (l_2 - 2)\ Y - X\ $
	BSWF	$Z(\lfloor (l_1 + l_2)/2 \rfloor)$	$Z(\lceil (l_1 + l_2)/2 \rceil)$	1	$l_1 + l_2 - 2$
(ii-i)	BSW	$X \cap Y(l_1), X - X \cap Y(\lceil l_1/2 \rceil), Y - X \cap Y(\lfloor l_2/2 \rfloor)$	$X \cap Y(l_2), X - X \cap Y(\lfloor l_1/2 \rfloor), Y - X \cap Y(\lceil l_2/2 \rceil)$	$\ X\  + \ Y\ $ $- 2\ X \cap Y\ $	$(l_1 + l_2 - 2)\ X \cap Y\  + (l_1 - 2)\ X - X \cap Y\ $ $+ (l_2 - 2)\ Y - X \cap Y\ $
	BSWF	$Z(\lfloor (l_1 + l_2)/2 \rfloor)$	$Z(\lceil (l_1 + l_2)/2 \rceil)$	2	$l_1 + l_2 - 2$
(ii-ii)	BSW	$X(1), Y - X \cap Y(\lfloor l_2/2 \rfloor)$	$X \cap Y(l_2), Y - X \cap Y(\lceil l_2/2 \rceil)$	$\ Y - X \cap Y\ $	$(l_2 - 1)\ X \cap Y\  + (l_2 - 2)\ Y - X \cap Y\ $
	BSWF	$Z(\lfloor (1 + l_2)/2 \rfloor)$	$Y(\lceil (1 + l_2)/2 \rceil)$	1	$l_2 - 1$
(ii-iii)	BSW	$X \cap Y(l_1), X - X \cap Y(\lceil l_1/2 \rceil)$	$Y(1), X - X \cap Y(\lfloor l_1/2 \rfloor)$	$\ X - X \cap Y\ $	$(l_1 - 1)\ X \cap Y\  + (l_1 - 2)\ X - X \cap Y\ $
	BSWF	$X(\lceil (l_1 + 1)/2 \rceil)$	$Z(\lfloor (l_1 + 1)/2 \rfloor)$	1	$l_1 - 1$
(iii-i)	BSW	$X(\lceil l_1/2 \rceil), Y(\lfloor l_2/2 \rfloor)$	$X(\lfloor l_1/2 \rfloor), Y(\lceil l_2/2 \rceil)$	$\ X\  + \ Y\ $	$(l_1 - 2)\ X\  + (l_2 - 2)\ Y\ $
	BSWF	$Z(\lfloor \max(l_1, l_2)/2 \rfloor)$	$Z(\lceil \max(l_1, l_2)/2 \rceil)$	2	$\max(l_1, l_2) - 2$
(iii-ii)	BSW	$X(1), Y(\lfloor l_2/2 \rfloor)$	$Y(\lceil l_2/2 \rceil)$	$\ Y\ $	$(l_2 - 2)\ Y\ $
	BSWF	$Z(\lfloor l_2/2 \rfloor)$	$Y(\lceil l_2/2 \rceil)$	1	$l_2 - 2$
(iii-iii)	BSW	$X(\lceil l_1/2 \rceil)$	$X(\lfloor l_1/2 \rfloor), Y(1)$	$\ X\ $	$(l_1 - 2)\ X\ $
	BSWF	$X(\lceil l_1/2 \rceil)$	$Z(\lfloor l_1/2 \rfloor)$	1	$l_1 - 2$

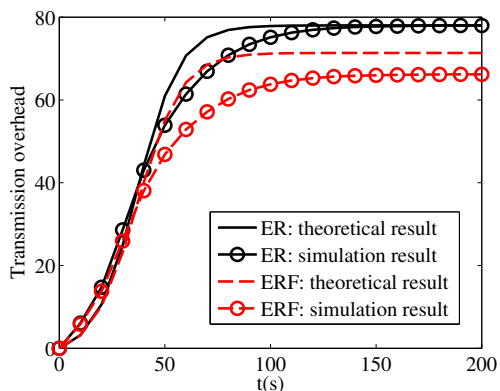


Fig. 7. Comparison of analytical and simulation results for the total transmission overhead of ER and ERF schemes;  $N=40$  nodes exist in a  $600 \times 600$  network, and  $K = 2$ ; All nodes move according to the Random Way Point model.

## APPENDIX G

### COMPARISON BETWEEN BSW AND BSWF

According to Definition 6 and Definition 7, Table 6 lists the packets and forwarding tokens carried by  $u_1$  and  $u_2$  after they encounter and perform data forwarding under various rules by using BSW and BSWF, respectively. Rule (iv) is not listed since both  $u_1$  and  $u_2$  remain unchanged. We also list the transmission overheads of the two schemes, including current overhead and potential overhead. Current overhead means the required number of transmissions by the current forwarding process between  $u_1$  and  $u_2$ . For example,  $u_2$  forwards a copy of each packet in  $Y - X$  to  $u_1$  under Rule (i) by using BSW, so the current transmission overhead is  $\|Y - X\|^\dagger$ ; by contrast,  $u_2$  only forwards a copy of the fused packet  $Y$  to  $u_1$  by using BSWF, so the current transmission

overhead is only 1. Potential overhead means that a node needs to spawn and forward multiple additional copies to other relay nodes according to the number of forwarding tokens it owns, unless it only has one forwarding token or it encounters the destination. For example,  $u_1$  needs to spawn and forward  $l_1 - 1$  copies of each packet in  $X$ , and  $\lfloor l_2/2 \rfloor - 1$  copies of each packet in  $Y - X$  to other relay nodes, and  $u_2$  needs to spawn and forward  $l_2 - 1$  copies of each packet in  $X$ , and  $\lceil l_2/2 \rceil - 1$  copies of each packet in  $Y - X$  to other relay nodes, under Rule (i) by using BSW, so the potential transmission overhead is  $(l_1 + l_2 - 2)\|X\| + (l_2 - 2)\|Y - X\|$ ; by contrast,  $u_1$  needs to spawn and forward  $\lfloor (l_1 + l_2)/2 \rfloor - 1$  copies of  $Z$  to other relay nodes, and  $u_2$  needs to spawn and forward  $\lceil (l_1 + l_2)/2 \rceil - 1$  copies of  $Z$  to other relay nodes by using BSWF, so the potential transmission overhead is  $l_1 + l_2 - 2$ .

## APPENDIX H

### PROOF OF THEOREM 4

From Table 6, we know that both the current and potential overheads of BSWF are lower than or equal to that of BSW under each rule. Thus, the total transmission overhead of BSWF is lower than or equal to that of BSW.

## APPENDIX I

### PROOF OF THEOREM 5

From Table 6, we know that the data fusion can increase the propagation speed of original packets compared to that without data fusion. Let's take Rule (i) for example. By using BSW,  $u_1$  and  $u_2$  will have  $l_1 + l_2$  forwarding tokens for each original packet in  $X$ , and  $l_2$  forwarding tokens for each packet in  $Y - X$  altogether. By contrast,

$\dagger$ .  $\|\cdot\|$  denotes the cardinality of a set, namely the number of original packets in the set.

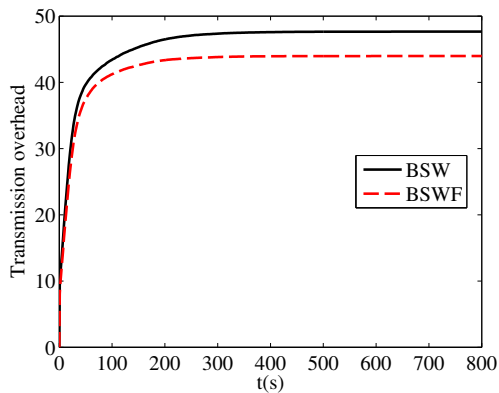


Fig. 8. Comparison of simulation results for the transmission overhead of BSW and BSWF schemes;  $N=40$  nodes exist in a  $600 \times 600$  network, and  $K = 10, L = 5$ ; All nodes move according to the Random Way Point model. A static sink is placed in the center of the simulation area. The average results are obtained by 1000 runs of simulations.

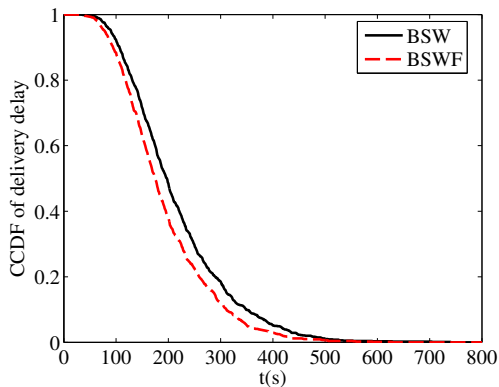


Fig. 9. Comparison of simulation results for the delivery delay of BSW and BSWF schemes. The simulation environments are the same as in Fig. 8.

by using BSWF,  $u_1$  and  $u_2$  will have  $l_1 + l_2$  forwarding tokens for the fused packet  $Z = f(X, Y) = Y$  altogether. It means that  $u_1$  and  $u_2$  will have  $l_1 + l_2$  forwarding tokens for each original packet in  $X$  and  $Y - X$  altogether in essence. It also means that the propagation speed of each original packet in  $Y - X$  will be increased if  $l_1 > 0$ . Only when  $l_1 = 0$ , the propagation speed will remain unchanged. For other rules, we can also obtain the same conclusion. Thus, the expected delivery delay of BSWF is lower than that of BSW.

## APPENDIX J

### SIMULATION RESULTS ON BSW AND BSWF SCHEMES UNDER THE RANDOM WAY POINT MODEL

In Fig. 8, we compare simulation results for the transmission overhead of BSW and BSWF schemes varying with the time. We observe that the transmission overhead of the BSW scheme is lower than  $K \times L = 50$ , and the transmission overhead of the BSWF scheme is 7.74%

lower than that of the BSW scheme when  $t = 800s$ . Fig. 9 shows simulation results for the CCDF of delivery delay of the two schemes. We observe that the BSWF scheme has lower delivery delay than the BSW scheme. The average delivery delay of the BSWF scheme is 192.05s, which is 10.71% than that of the BSW scheme (215.09s).