A Novel Delay-Aware and Privacy-Preserving Data-Forwarding Scheme for Urban Sensing Network

Di Tang and Jian Ren, *Senior Member, IEEE*

***Abstract*—Security, communication delay, and delivery ratio are essential design issues in urban sensing networks. To address these three issues concurrently, we propose a novel delay-aware privacy-preserving (DAPP) transmission scheme based on a combination of two-phase forwarding and secret sharing. In DAPP, the collected data are first split into pieces, and then, each piece is relayed to an application data server by randomly selected intermediate delivery nodes. The two-phase forwarding method detaches the connection between the application data server and the source node, which renders it infeasible for the application data server to estimate the source node identity. The underlying secret sharing scheme and dynamic pseudonym ensure confidentiality of the collected data and anonymity of participating users. Through DAPP, we can also verify data integrity in hostile networks. Moreover, DAPP provides a framework to achieve a design tradeoff among security, communication delay, and delivery ratio. The security analysis demonstrates that DAPP can preserve location privacy while defending against side information attack. Our theoretical analysis and numerical results show that the three design issues can be adjusted to meet various security and practical implementation goals.**

***Index Terms*—Delay aware, mobile service, privacy preserving, urban sensing.**

I. Introduction

RECENT technological advances make urban sensing networks technically and economically feasible to be widely used in civilian applications, such as monitoring of urban environment, traffic condition, and patient care. Urban sensing, which is also known as participatory sensing, relies on sensors embedded in human-carried mobile devices or vehicular electrical devices to collect and report the sensing data. The ubiquitous electrical devices carried by human and vehicles are gradually replacing the traditional static sensor networks for many urban area applications.

Participatory sensing networks differ significantly from traditional wireless sensor networks. First, participatory sensor nodes are embedded in rechargeable mobile devices, such as smartphones, iPads, laptops, and auto computers. The powerful resources equipped in these devices enhance their capabilities in computation, sensing, data storage, reliable data communication, and energy lifetime dramatically. Hence, energy efficiency is no longer as critical as in traditional wireless sensor networks. Second, in urban sensing networks, sensing data collection and reporting no longer rely on fixed infrastructure. The data can be forwarded by mobile nodes to the sink either directly or indirectly in multiple hops. Due to mobility, network topology structure constantly changes, which makes end-to-end communication delay, message delivery ratio, and quality of service (QoS) some of the most essential performance evaluation metrics in urban sensing networks. Third, these devices are owned and operated by individuals. Instead of being only data consumers, these devices could also collect sensed data. It makes data collection more directly related to people's lives; hence, privacy becomes one of the major concerns for participating users. In fact, the data collecting and reporting services put users' privacy in jeopardy since both the delivery nodes and the wireless access points (APs) are able to identify the data owner through direct communications. The uploaded data are collected with spatiotemporal information, which may be used to extract or infer sensitive and private information of the users. As a result, location information and side information may be correlated to recover the trajectory of participating users.

The unique nature of participatory sensing networks creates new security and performance requirements. In this paper, we propose a decentralized data-forwarding scheme to preserve trajectory privacy of participating users based on secret sharing and dynamic pseudonym. Delay-aware privacy preserving (DAPP) provides a design tradeoff framework to address security, communication delay, and delivery ratio based on the selection of an (k, n) secret sharing scheme. We provide quantitative analysis on security, communication delay, and delivery ratio based on the parameter settings.

Our contributions in this paper can be summarized as follows.

1) We propose a DAPP data-forwarding scheme to protect trajectory privacy of participant users in urban sensing networks.
2) We devise a secret-sharing-based secure message delivery option that can provide data integrity verification of the recovered data and maximize the message delivery ratio by introducing redundancy to the reported data.
3) We present a dynamic pseudonym scheme to defend against side information attacks.
4) The proposed scheme is able to achieve a tradeoff among security, communication delay, and delivery ratio through adjustable parameters.

Manuscript received October 31, 2014; revised February 3, 2015; accepted March 5, 2015. Date of publication March 20, 2015; date of current version April 14, 2016. This work was supported in part by the U.S. National Science Foundation under Grant CNS-0845812, Grant CNS-1117831, Grant CNS-1217206, and Grant ECCS-1232109. The review of this paper was coordinated by Prof. Y. Zhang.

The authors are with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824 USA (e-mail: ditony@msu.edu; renjian@msu.edu).

Digital Object Identifier 10.1109/TVT.2015.2415515

5) We quantitatively analyze performance of the proposed scheme on security, communication delay, and delivery ratio.

The remainder of this paper is structured as follows. In Section II, the related work is reviewed. The system model is described in Section III. The proposed scheme is presented in Section IV. In Section V, security analysis of the proposed scheme is conducted. Section VI provides performance analysis and numerical results. We conclude in Section VII.

II. Related Work

The participatory sensing concept was first proposed in [1]. It was further extended to urban sensing and people-centric sensing in [2]. The key idea for participatory sensing is that the network relies on people carried mobile devices to collect data sensed in urban area. An urban sensing network has also been introduced to vehicular sensing platform [3], [4] to monitor traffic conditions in urban areas. This network provides a flexible method to collect valuable environmental information. However, it also raises significant privacy concerns.

Location privacy for urban sensing networks was first discussed in [5]. The main idea is to blur location information based on tessellation and clustering against adversaries. The reported data are aggregated to enhance privacy protection [6], [7]. PriSense was designed based on slicing and mixing [7]. The sensed data are divided into pieces and randomly disseminated to cover nodes. At the cover nodes, the reported data are aggregated and reported to the server. This method can effectively preserve the identity privacy; however, it can only provide statistical sensing results for data users.

Mix zones with a pseudonym was proposed in [8] and [9]. To conceal real identities and accomplish k-anonymity within the mix zone, a trusted third party randomly assigns pseudonyms to the data reporters when they enter into or departure from the mix zone. Nevertheless, mix zones cannot be placed over the entire network to provide service for data collectors.

Spatial cloaking [10] was proposed to enable data collectors to adjust the resolution of location information along spatial or temporal dimensions to avoid exposure of user's exact location. This methodology was further studied in [11]–[14] to derive a tradeoff between QoS and anonymity. Unfortunately, the privacy protection is achieved by sacrificing the resolution of either the spatial or the temporal dimensions.

Dummy location was studied in [15]–[17]. The fundamental idea is to report dummy locations to conceal the actual location of the reported data. Suppression-based techniques [18] was also proposed to blur the reported locations by converting the database of trajectories. In these methods, the privacy can be protected by obscuring the exact locations, which may induce information loss for data service.

Encryption-based algorithms and data exchange schemes are proposed in [19] and [20]. In [19], the sensed data is encrypted and disseminated to replica sensors. The replica sensors then store the received data and relay it upon receiving inquiries. As a consequence, the source node identity can be concealed by the replica sensors. Nevertheless, the shared key may either help the operator to identify the source node or enable the malicious replica nodes to decrypt the data readings. In [20], the collected sensing readings are exchanged between participants within the wireless communication range. The data can be exchanged multiple times to prevent adversaries from correlating the data and its identity. Hu and Shahabi [21] proposed to forward the collected data from friends to friends of the source user until the required number of hops has been reached. In these schemes, source node may be directly identified by intermediate nodes through wireless communications. Additionally, malicious intermediate nodes can also reveal and tamper with data contents through the forwarding procedure. There is little flexibility in delivery ratio when the data are being dropped or tampered with.

III. Models and Assumptions

A. System Model

The urban sensing network relies on a set of sensor nodes embedded in mobile devices or vehicular systems for data sensing and reporting. These devices can get wireless Internet access intermittently through wireless APs in the sensing area. The APs, such as Wi-Fi APs, may be owned and operated by the government, organizations, or individuals. They are directly connected to the application data server. In this network, the surrounding environment information is collected by the participating mobile devices and eventually sent to the application data server.

The application data server provides environmental sensing services for data consumers and disseminates the requested tasks to mobile devices carried by the participants. Mobile devices can join the system at will to participate in data sensing. They are required to report the collected data to the application data server once they accept the tasks. To provide precise services to data consumers, the data are required to be collected with spatiotemporal information. Here, we may use the term "report location" to indicate the spatiotemporal information. The data format is $\{_p\mathsf{ID}, (\hat{x}, \hat{y}), \hat{T}_d\}$, where $_p\mathsf{ID}$ is the pseudonym of the data source, $(\hat{x}, \hat{y})$ is the coordinator where data is collected, and $\hat{T}_d$ is the collecting time. This information may divulge details about the participating user's location. Due to characteristics of opportunistic sensing and physical infrastructure of the network, we only deal with sensing data that are delay tolerant.

B. Adversary Model

Our adversary model is similar to [6]. It can be summarized as follows.

- The adversary can be any parties in the network, including individual sensing nodes, wireless APs, and even the application data servers.
- Adversaries are generally assumed honest but curious. We also assume that they may drop or tamper with the reported data due to their own interests.
- The collected data are eventually forwarded to the application data server, which enables it to disclose their reported location. The application data server may try to uncover source identities of the received data and the trajectory.

- The participating user list is kept secret to delivery nodes and public; however, the application data server can access the list due to its administrative right.
- Intermediate delivery nodes can obtain the source pseudonym of the received data. They may collude to obtain the location privacy information or forge collected data to fool the application data server.

C. Side Information Attack

Side information attack includes both direct side information attack and indirect side information attack. The *direct side information* refers to the information that an adversary can acquire based on direct access to the communication. In our case, the direct side information may include information $\{{}_p\mathsf{ID}, (\hat{x}, \hat{y}), \hat{T}_d\}$ of a particular event. In particular, the adversary can record the set of identities $\mathscr{I}$ that appear within its communication range. The nature of wireless communications makes adversary able to extrapolate the source node located within its communication range. If $\mathscr{I}$ is small enough by combining direct side information obtained, the adversary may derive the actual ID and correlate it to the ${}_p\mathsf{ID}$.

The *indirect side information* is defined as the information that an adversary can obtain through indirect channels, such as media, video, and web blog published on the Internet, of a particular event. In the DAPP scheme, an adversary may receive data d, which contains $\{{}_p\mathsf{ID}, (\hat{x}, \hat{y}), \hat{T}_d\}$. To derive the actual source identity ID of ${}_p\mathsf{ID}$, it may search through the public domain to find the people who have visited location $(\hat{x}, \hat{y})$ at time $\hat{T}_d$. In this way, the adversary may either completely identify the actual ID or limit it to a smaller subset.

D. Mobility Model

In urban sensing networks, sensor nodes are embedded in the mobile devices carried by people or vehicles. In this paper, we apply the Manhattan street pedestrian model described in [22]. The analysis for vehicular networks is similar to the mobility model presented in [23]. The Manhattan street model is proposed to emulate the movement pattern of pedestrian on the streets. In this model, sensor nodes move on a two-way street segment. The street segment can be modeled as a real street between two intersections. The arrivals and departures of mobile nodes occur at the endpoints of the street segment. The velocities of the mobile sensor nodes are independent and identically distributed (i.i.d.) random variables with a probability density function $f_v(v)$. The direction and speed of a node remain constant in one segment. Participating users arriving at both endpoints can be modeled as a Poisson Process [24]. The total arrival rate is denoted as λ. At each endpoint, the mobile node may alter its direction with a predetermined probability. The four directions at a crossroad are expressed as f(orward), l(eft), r(ight) and b(ackward), respectively. The probability for going these directions at the intersection are P_f, P_r, P_l, P_b. This model assumes that communications between two nodes in different street segments is not possible. As a result, an existing connection breaks at the endpoints.

E. Design Goal

Our design goals can be described as follows.

- Adversaries should not be able to obtain real identities of participating users without side information.
- The malicious delivery node should not be able to discover spatiotemporal information of the reported data.
- Adversaries should not be able to recover data trajectory by collecting side information.
- The proposed scheme should provide data integrity verification for the recovered data.
- The proposed scheme should be able to provide high delivery ratio in case some data pieces are dropped or tampered with by malicious delivery nodes.
- The proposed scheme should provide flexibility and diversity in protocol design.

IV. The Proposed Delay-Aware-Privacy-Preserving Scheme

Here, we present a novel DAPP data reporting scheme based on a combination of Shamir's secret sharing [25] and two-phase routing. It can ensure data confidentiality and provide a data integrity verification option for the reported data. We also propose a dynamic pseudonym scheme to guarantee anonymity of the source node.

A. Overview of the Proposed Privacy Scheme

To transmit collected data, the source node generates n data pieces from the collected data based on the Shamir's secret sharing scheme. It then generates a unique pseudonym for each data piece as its identity to conceal the source information. The data pieces are then forwarded to n randomly selected participating users, named as delivery nodes, within the communication range. Delivery nodes relay the received data pieces to the application data server through nearby APs. Upon receiving k or more data pieces, the application data server is able to reconstruct the original collected data. To ensure integrity of the recovered data, both the original data and its hash value will be transmitted together.

B. Secret Sharing

In Shamir's (k, n) secret sharing scheme, to share a secret S, the secret holder picks a random $k-1$ degree polynomial $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{k-1}x^{k-1}$ over a finite field $\mathbb{Z}_q$, where $a_0 = S$ and $q > \max(S, n)$ is a prime number. $\mathcal{H}$ generates n data pieces $S_i = (i, f(i)), i = 1, \ldots, n$. The secret holder distributes the n data pieces to n individual users.

The secret S can be recovered by k or more available data components using Barycentric Lagrange Interpolation. The computational complexity to recover the secret S is only $\mathcal{O}(n)$. More specifically, for k available data components, S can be recovered as follows.

1) For any k out of n secret pieces $(i, f(i))$, without loss of generality, we assume $i = 0, \ldots, k-1$. Define $l(x) = (x - x_0)(x - x_1) \cdots (x - x_{k-1})$.

2) Compute the weight

$$w_j = \frac{1}{\prod_{i=0, i\neq j}^{k}(x_j - x_i)}. \quad (1)$$

3) Let

$$l_j(x) = l(x)\frac{w_j}{x - x_j},$$

$$L(x) = l(x)\sum_{j=0}^{k}\frac{w_j}{x - x_j}y_j. \quad (2)$$

4) The shared secret is $S = L(0)$.

C. Dynamic Pseudonyms

Pseudonyms have been widely used to prevent adversaries from obtaining source identities. However, adversaries may use side information to link multiple pseudonyms or correlate the pseudonyms to the actual identity. To solve this problem, we propose a new method to generate dynamic pseudonyms using an ID hash chain. Instead of simply using hash functions, we enable the source node to employ a secret key in the hash chain.

Suppose ID is the real identity of a participating user. To conceal it, the user replaces ID with a dynamic changing pseudonym ${}_p\mathsf{ID}_i$ for each message transmission, where i is related to the order of the message. Each pseudonym is generated using a one-way hash function $H(\cdot)$ based on the previous pseudonym and the secret key. The ID hash chain $\{{}_p\mathsf{ID}_1, {}_p\mathsf{ID}_2, \ldots\}$ for the participating user ID is generated as follows:

$${}_p\mathsf{ID}_1 = H(\mathsf{ID}, K),$$

$${}_p\mathsf{ID}_2 = H({}_p\mathsf{ID}_1, K),$$

$$\ldots$$

where K is a secret key of the message source. Without knowing the secret key and the real identity, it is infeasible to establish a linkage between pseudonyms or between a pseudonym and the real identity.

D. DAPP Scheme

In DAPP, to ensure security of the reporting data d, we generate n data pieces based on Shamir's secret sharing scheme. Each data piece is then forwarded to a randomly selected delivery node. In this way, the original data can be concealed among n trustworthy secret holders. The shared data d can be recovered by any k or more data pieces. Since we assume the delivery nodes may tamper with the reporting data, the application server may receive incorrect data pieces. Therefore, the application data server may not be able to recover the original data. To deal with this issue, the proposed scheme is designed to be able to verify integrity of the recovered data.

We assume the source node $\mathcal{H}$ has data d to transmit. It first computes the hash value $H(d)$, where $H(\cdot)$ is a one-way hash function. d and $H(d)$ are treated as two independent secret data pieces. We randomly select two (k, n) secret sharing schemes to share d and $H(d)$ separately. The polynomial computation is operated over $\mathbb{Z}_q$. The procedure is described as follows.

1) $\mathcal{H}$ chooses a prime $q > \max(d, n)$.
2) $\mathcal{H}$ constructs two coefficient sets, i.e., $\mathcal{A}$ and $\mathcal{B}$. Each set contains $k-1$ randomly selected coefficients, i.e., $\mathcal{A} = \{a_0 = d, a_1, \ldots, a_{k-1}\}$ and $\mathcal{B} = \{b_0 = H(d), b_1, \ldots, b_{k-1}\}$ from $\mathbb{Z}_q$.
3) $\mathcal{H}$ generates two random polynomials over $\mathbb{Z}_q$ as follows:

$$f_a(x) = \sum_{i=0}^{k-1} a_i x^i, \quad (3)$$

$$f_b(y) = \sum_{i=0}^{k-1} b_i y^i. \quad (4)$$

4) $\mathcal{H}$ computes n components $f_a(i), f_b(i)$ from d and $H(d)$, respectively, $i = 1, \ldots, n$.
5) The ith data piece for d and $H(d)$ is $(i, f_a(i), f_b(i))$, $i = 1, \ldots, n$.

The proposed data distribution scheme is summarized in Algorithm 1.

Algorithm 1 Data Distribution

1: Choose a prime $q > \max(d, n)$.
2: Construct two coefficient sets $\mathcal{A}$ and $\mathcal{B}$ from $\mathbb{Z}_q$ randomly.
3: Based on $\mathcal{A}$ and $\mathcal{B}$, construct two polynomials $f_a(x) = \sum_{i=0}^{k-1} a_i x^i$, and $f_b(y) = \sum_{i=0}^{k-1} b_i y^i$ over $\mathbb{Z}_q$ to share d and $H(d)$, respectively, as two secret values.
4: Compute n data pieces $d_i = (i, f_a(i), f_b(i), q), i = 1, \ldots, n$.
5: Generate a pseudonym ${}_p\mathsf{ID}$ for d using ID hash chain.
6: **for** each $i \in [1, n]$ **do**
7: Send d_i to a randomly selected neighboring node M_i.
8: Insert an interval t_α before sending d_{i+1}.
9: The neighboring node M_i forwards d_i to a wireless access point.
10: **end for**

DAPP includes two phases in data forwarding. In the first phase, the generated n data pieces are distributed to n randomly selected delivery nodes before being relayed to the application server. The data source may choose to add a time interval t_α between transmissions of two consecutive data pieces. A properly selected t_α can effectively control the probability for any single delivery node to receive multiple data pieces, particularly in sparse networks. Since delivery nodes may not be completely trusted, the integrity of the reconstructed data has to be verified. The reconstruction algorithm follows the Barycentric Lagrange Interpolation defined in equation (1) and equation (2). It is described in Algorithm 2.

Algorithm 2 Data Reconstruction and Verification

1: Suppose the server has received at least k out of n data pieces $(i, f_a(i), f_b(i))$ for data $(d, H(d))$.
2: The application data server reconstructs the original data using the Barycentric Lagrange Interpolation algorithm described in equation (1) and equation (2).
3: The application data server verifies integrity of the reconstructed data by checking whether $D = H(d)$ holds true.

V. Security Analysis

Source privacy information can be specified by two equally essential parts: spatiotemporal information and identity information. Only when both are exposed that the complete privacy information is disclosed.

Here, we will provide quantitative analysis that DAPP can provide both identity and spatiotemporal information from being disclosed to adversaries. We first introduce several definitions and metrics.

A. Definitions and Security Metrics

Definition 1 (Identity Information Leakage): The identity information leakage for data d is defined as the probability that an adversary is able to derive the source identity d_I of data d. We denote it as $P(d_I)$.

Definition 2 (Location Information Leakage): The location information leakage for data d is defined as the probability that an adversary is able to derive the spatiotemporal information d_L of data d. The probability is denoted $P(d_L)$.

To measure identity information loss of the received data, we introduce mutual information.

Definition 3 (Identity Information Loss): Let X be the set of possible identities estimated by adversaries prior to receiving a message, and Y be the set of estimated source identities after receiving the message. The identity information loss for forwarding the message event is defined as the mutual information between the X and Y. That is

$$I(X;Y) = H(Y) - H(Y|X). \tag{5}$$

Notice that source privacy information consists of both spatiotemporal information and its correlated identity. We introduce the following metric to measure the joint information loss.

Definition 4 (Joint Information Leakage): The joint identity and location privacy information leakage P_T for data d is defined as the joint probability that the adversary obtains the spatiotemporal information and the identity of data d:

$$P_T(d) = P(d_L, d_I) = P(d_L|d_I)P(d_I) = P(d_I|d_L)P(d_L) \tag{6}$$

where d_I and d_L denote the identity and location information of data d, respectively.

B. Identity Information Loss

The identity information loss can be derived by the following theorem.

Theorem 1: Assume that an adversary is able to limit the message source node to a potential subset $Y \subset X$ after receiving a message. Then, the identity information loss is

$$I(X;Y) = \log|X| - \log|Y|, \tag{7}$$

where $|X|, |Y|$ denote the cardinalities of X and Y, respectively.

Proof: In set X, if each node can be the source node with equal probability, then the probability for each node in the set to be selected as the source node is $1/|X|$. After a message is received, the adversary may limit the message source node to set Y. If each node in set Y can be the source node with equal probability, the probability for each node in Y to be selected as the source node is $1/|Y|$. The identity information loss can be calculated by the following:

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= -\sum p(x)\log p(x) + \sum_{y\in Y} p(y) \sum p(x|y)\log p(x|y) \\ &= -|X| \cdot \frac{1}{|X|}\log\frac{1}{|X|} + \sum_{y\in Y} p(y)\left(|Y| \cdot \frac{1}{|Y|}\log\frac{1}{|Y|}\right) \\ &= \log|X| - \log|Y|. \end{aligned}$$

■

1) Without Side Information Attack: For security attack without side information, based on Definitions 1 and 3 and Theorem 1, we have the following corollary.

Corollary 1: The proposed DAPP scheme can achieve unconditional message source privacy protection for participating users under message-ID-based attack. That is the set of possible source node X prior to data transmission is the same as the set of the possible source node Y after message transmission. As a result, we have $I(X;Y) = 0$.

Proof: DAPP introduces a two-phase message forwarding: distribution of data pieces to randomly selected delivery nodes and message forwarding to the application server through the APs. Assume an adversary receives a data piece d_i with a unique pseudonym ${}_p\mathsf{ID}$. Prior to receiving d_i, the potential set for data sources can be arbitrary nodes in the entire set of the population in the urban area of the delivery nodes since they are not entitled to access the list of the participating users. For the application data server, since it has access to the participating user list, the priori estimated set is the same as the participating users.

Through message forwarding, the adversary obtains ${}_p\mathsf{ID}$ instead of ID of d_i. The proposed dynamic pseudonym makes it infeasible for adversaries to derive ID from ${}_p\mathsf{ID}$ without secret key K. Hence, the posterior estimated identity set Y based on message forwarding event is also equal to X. Therefore, $I(X;Y) = 0$. ■

2) Side Information Attack: Side information attack is envisioned as one of the major threats to data privacy. It may divulge part or even the entire source identity information.

Corollary 2: For DAPP, through direct side information attack, the identity information loss to the delivery node $(\widetilde{I}^D)$,

the AP ($\widetilde{I}^S$) and the application data server ($\widetilde{I}^A$) can be derived as follows:

$$\widetilde{I}^D(X;Y) = \log|X| - \log|Y|, \tag{8}$$

$$\widetilde{I}^S(X;Y) = \widetilde{I}^A(X;Y) = 0, \tag{9}$$

where $|X|, |Y|$ denote the cardinalities of X and Y, respectively.

Proof: Direct side information is obtained through traffic monitoring during message forwarding. The data pieces are relayed to the application data server by randomly selected delivery nodes. The application data server and APs may possibly collect direct side information only about delivery nodes rather than the source node. Therefore, they are unable to reduce the size of X. As a consequence, $X = Y$. Therefore, we have

$$\widetilde{I}^S(X;Y) = \widetilde{I}^A(X;Y) = 0. \tag{10}$$

The potential event set X is the same as the total population in the network. Since the delivery node can acquire direct side information about the source node, after the message is being transmitted, the delivery node may be able to limit the message source to a subset Y, which is equal to the population in the communication range of the source. Based on Theorem 1, we have

$$\widetilde{I}^D(X;Y) = \log|X| - \log|Y|. \quad \blacksquare$$

Corollary 3: Through indirect side information attack, the identity information loss to the delivery node, the AP, and the application data server can be derived as follows:

$$\bar{I}^S(X;Y) = \log|X| - \log|Y| \tag{11}$$

$$\bar{I}^D(X;Y) = \bar{I}^A(X;Y) = 0. \tag{12}$$

Proof: The indirect side information can be used to derive the source identity by exploring the matched spatiotemporal information through public information. The key procedure of this attack lies in the recovery of the reported location $\{(\hat{x}, \hat{y}), T_d\}$ in d. However, DAPP applies secret sharing to ensure confidentiality, which makes it infeasible to reveal the data content by single data piece. Thus, neither an AP nor a delivery node can reveal the source identity of the received data piece. The identity information loss to them is

$$\bar{I}^D(X;Y) = \bar{I}^A(X;Y) = 0. \tag{13}$$

In addition, data pieces are eventually sent to the application data server, which enables it to recover the reported location. Due to access right, application data server can acquire the participating user list; thus, the estimated identity set X is equal to the list of participating users. Based on indirect side information, it can limit X into a subset Y, which depends on the accuracy of indirect side information, i.e.,

$$\bar{I}^S(X;Y) = \log|X| - \log|Y|. \quad \blacksquare$$

C. Location Information Leakage

Spatiotemporal information is equally essential for source privacy. In the subsequent analysis, we will discuss the location information leakage of DAPP. We first prove that the data distribution process follows the Poisson process.

Theorem 2: The process of data distribution in the first forwarding phase of the proposed DAPP scheme is a Poisson process.

Proof: See the Appendix. ■

1) Location Information Leakage to Individual Delivery Node: In this paper, the Manhattan street model is applied to analyze the location information leakage to a malicious delivery node. In this model, we denote ℓ as the length of a street segment. Assume that S is the source node and M is a malicious node. v_S and v_M are the velocities of S and M, respectively. Denote the probability density function for v as $f(v)$. Let t_M be the time duration that a malicious node stays in the source node communication range Δ. Then, we have $t_M = 2\Delta/z$, where $z = |v_M - v_S|$. We also use $P(t_M > kt_\alpha)$ to denote the probability that a sensor node stays in the source node communication range at least kt_α seconds. Let P_M be the probability that a malicious node M is selected as the delivery node in one data piece distribution, and let c be the number of times that M has been selected as the delivery node.

Theorem 3: Assume velocities of the malicious node and the source node are two i.i.d. random variables, then the location information leakage to the malicious delivery node can be derived as follows:

$$P^D(d_L) = \int_{-\frac{2\Delta}{kt_\alpha}}^{\frac{2\Delta}{kt_\alpha}} f(z)\,dz \times \sum_{i=k}^{n}\left[\left(\frac{1}{4}\right)^{\left\lfloor\frac{2it_\alpha v_S}{\ell}\right\rfloor}\binom{n}{i}P_M^i(1-P_M)^{n-i}\right] \tag{14}$$

where $f(z)$ is the joint probability density function of $z = |v_M - v_S|$.

Proof: Since the data content can be closely related to the reporting location, the leakage of message content may inadvertently expose the source location. To deal with this potential security issue, DAPP utilizes secret sharing to ensure data confidentiality by generating multiple seemingly meaningless data pieces and delivers each piece through a randomly selected delivery node. In this way, the malicious node will not be able to get the content of the data unless it is capable of collecting up to k or more data pieces.

There are two possible scenarios for a malicious node to obtain k data pieces: 1) The malicious node comes across the data source node at least k times to collect the data pieces from the source node during the entire duration of the message transmission; and 2) the malicious node stays in the communication range of the source node at least kt_α seconds.

In the first case, since velocities of the participating users are independent, the spatial dependence degree of two arbitrary nodes is approximately 0 according to [26], which means that the probability for two nodes to meet for the second time is approximate to zero within a limited short time duration. Thus, probability for the first case can be viewed as 0.

The success of the malicious node is determined by the likelihood of the second case. If we assume that $f(z)$ is the joint probability density function of $z = |v_M - v_S|$, then we have

$$\begin{aligned} P(t_M \geq kt_\alpha) &= P\left(\frac{2\Delta}{z} \geq kt_\alpha\right) \\ &= P\left(z \leq \frac{2\Delta}{kt_\alpha}\right) \\ &= \int_{-\frac{2\Delta}{kt_\alpha}}^{\frac{2\Delta}{kt_\alpha}} f(z)\,dz. \end{aligned} \tag{15}$$

The success of case 2 requires the malicious node to be selected as the delivery node at least k times. In DAPP, the source node selects delivery nodes based on the principle of randomness; thus, each node in the source communication range can be selected as delivery node with equal probability $P_M = 1/|X|$, where $|X|$ is the total number of participants in the communication range of S. Based on Theorem 2, the process of the participating users' arrival and departure can be modeled as $M/G/\infty$ based on *theory of queues*. In this model, the source node can be denoted as the service node. t_M is actually the service time for a given participating user. Moreover, suppose λ is the arrival rate and T is the average service time, we can have

$$|X| = \lambda \cdot T. \tag{16}$$

Since ℓ is the length of a street segment, we have $t_M \leq \ell/(\max[v_S, v_M])$. Without loss of generality, we may assume $v_S \geq v_M$, then

$$T = 2 \cdot \int_{\frac{2\Delta v_S}{\ell}}^{\nu} \frac{2\Delta}{z} f(z)\,dz, \tag{17}$$

where $\nu = V_{\max} - V_{\min}, V_{\max}$ and $V_{\min}$ represent the maximum and minimum speed of the participating users, respectively.

The location information leakage in one street segment can be described as the joint probability of $t_M \geq kt_\alpha$ and $c \geq k$, i.e.,

$$\begin{aligned} P^D(d_L) &= P(t_M \geq kt_\alpha, c \geq k) \\ &= P(t_M \geq kt_\alpha) P(c \geq k \,|\, t_M \geq kt_\alpha) \\ &= P(t_M \geq kt_\alpha) \sum_{i=k}^{n} \binom{n}{i} P_M^i (1 - P_M)^{n-i}. \end{aligned} \tag{18}$$

In (18), all data pieces are assumed to be distributed in one street segment. In fact, these data pieces may be distributed in several streets. S and M may change their velocities at the endpoint of the street segment. Hence, M can stay in the communication range of S only if they move in the same direction. To simplify the analysis, we assume that the probability for each direction is equal to $1/4$. Thus, the probability that S and M choose the same direction at an endpoint is equal to $(1/4)^2$. Therefore

$$\begin{aligned} P^D(d_L) &= P(t_M > kt_\alpha) \\ &\quad \times \sum_{i=k}^{n} \left[\left(\frac{1}{4}\right)^{\left\lfloor \frac{2it_\alpha v_S}{\ell} \right\rfloor} \binom{n}{i} P_M^i (1 - P_M)^{n-i} \right]. \end{aligned} \tag{19}$$

Combining (15) and (19), we can get (14). ■

From Theorem 3, we can derive the following corollary.

Corollary 4:

$$\lim_{kt_\alpha \to \infty} P^D(d_L) = 0. \tag{20}$$

Proof: From (14), we have

$$\begin{aligned} P^D(d_L) &= \int_{-\frac{2\Delta}{kt_\alpha}}^{\frac{2\Delta}{kt_\alpha}} f(z)\,dz \\ &\quad \times \sum_{i=k}^{n} \left[\left(\frac{1}{4}\right)^{\left\lfloor \frac{2it_\alpha v_0}{\ell} \right\rfloor} \binom{n}{i} P_M^i (1 - P_M)^{n-i} \right] \\ &\leq \int_{-\frac{2\Delta}{kt_\alpha}}^{\frac{2\Delta}{kt_\alpha}} f(z)dz \cdot \left(\frac{1}{4}\right)^{\left\lfloor \frac{2kt_\alpha v_S}{\ell} \right\rfloor}. \end{aligned} \tag{21}$$

Since both $\int_{-\frac{2\Delta}{kt_\alpha}}^{\frac{2\Delta}{kt_\alpha}} f(z)\,dz$ and $(1/4)^{\lfloor 2kt_\alpha v_S/\ell \rfloor}$ are asymptotically equivalent to zero as $kt_\alpha \to \infty$. Therefore, we have $\lim_{kt_\alpha \to \infty} P^D(d_L) = 0$. ■

2) Location Information Leakage to APs: In the second forwarding phase, delivery nodes transmit the data pieces to APs, which are assumed uniformly distributed in urban area. Thus, each AP has equal probability to receive the data piece, and we have the following corollary.

Corollary 5: The location information leakage to each AP can be computed by the following equation:

$$P^A(d_L) = \sum_{i=k}^{n} \binom{n}{i} \left(\frac{1}{\eta}\right)^i \left(1 - \frac{1}{\eta}\right)^{\eta - i}, \tag{22}$$

where η is the number of wireless APs in urban area. From this equation, we can see that the location information leakage to each delivery node and each AP is asymptotically equivalent to zero as $k \to \infty$.

D. Joint Identity and Location Privacy Information Leakage

We have discussed the identity information loss and the location information leakage separately. The results show that the delivery node may only obtain partial identity information that is insufficient to identify the data source. However, the reported location and the identity information may be dependent on side information attack. To elaborate privacy information loss, we further investigate privacy leakage when these two pieces are considered jointly.

TABLE I
JOINT PRIVACY INFORMATION LEAKAGE FOR COLLUSION ATTACKS, WHERE $n = 30$

Collusion $P_C(d_L)$ (%)	$k = 5$	$k = 10$	$k = 15$	$k = 20$	$k = 25$	$k = 30$
1%	1.16×10^{-3}	2.50×10^{-11}	1.35×10^{-20}	2.73×10^{-31}	1.36×10^{-43}	1.36×10^{-58}
5%	1.56	1.16×10^{-4}	2.31×10^{-10}	1.76×10^{-17}	3.32×10^{-26}	9.31×10^{-38}
10%	17.5	4.54×10^{-2}	3.56×10^{-6}	1.11×10^{-11}	8.60×10^{-19}	1.02×10^{-28}
15%	47.6	9.66×10^{-1}	7.08×10^{-4}	2.14×10^{-8}	1.65×10^{-14}	1.92×10^{-23}

Theorem 4: For the joint identity and location privacy information leakage, we have

$$\lim_{kt_\alpha \to \infty} P_T^D = 0, \tag{23}$$

$$\lim_{\substack{k=n \\ n\to\infty}} P_T^A = 0. \tag{24}$$

Proof: Based on Definition 4, we have

$$P_T^D \le P^D(d_L),\ P_T^A \le P^A(d_L).$$

The equalities in the given equations may hold if the adversary can collect k or more data pieces through side information attack. Thus, the joint identity and location privacy information leakages to an AP and a delivery node are both dependent on the parameters $(k, n; t_\alpha)$. Based on Corollary 4, we have

$$\lim_{kt_\alpha \to \infty} P_T^D \le \lim_{kt_\alpha \to \infty} P^D(d_L) = 0.$$

From (22), the joint identity and location privacy information leakage to an AP can be derived as follows:

$$0 \le \lim_{\substack{k=n \\ n\to\infty}} P_T^A \le \lim_{\substack{k=n \\ n\to\infty}} P^A(d_L) = \lim_{k\to\infty} \left(\frac{1}{\eta}\right)^k = 0. \quad \blacksquare$$

Similarly, for the application server, we have the following corollary.

Corollary 6: For the application data server, the joint identity and location privacy information leakage under side information attacks can be calculated by $P_T^S = 1/|Y|$.

E. Participating Nodes Collusion

Participating users may conspire or share information to infer privacy information of others. Assume there are $\kappa(\kappa \le k_1)$ malicious nodes in the network, where k_1 is the number of delivery nodes in urban sensing networks. The probability for an arbitrary delivery node to be malicious is κ/k_1. Hence, the location information leakage of the reported data d is

$$P^C(d_L) = \sum_{i=k}^{n} \binom{n}{i} \left(\frac{\kappa}{k_1}\right)^i \left(1 - \frac{\kappa}{k_1}\right)^{n-i}. \tag{25}$$

For each data piece d_i, the malicious delivery node may record a potential identity set through traffic monitoring, $\mathscr{I}^i = \{a_0^i, \ldots, a_l^i\}$. As the direct side information can be shared, the colluding users could limit the source identity to be an element in the intersection $\bigcap_{i=0}^{k} \mathscr{I}^i$. The joint identity and location privacy information $\bar{P}_T^C$ for collusion attacks falls as the location information leakage $P^C(d_L)$ decreases. Based on (25), $P^C(d_L)$ decreases to $(\kappa/k_1)^k$ as k increases to n. Then, $\bar{P}_T^C$ can be asymptotically equivalent to zero as $k = n$ and $k \to \infty$. Therefore, we have the following corollary.

Corollary 7:

$$\lim_{\substack{k=n \\ n\to\infty}} \bar{P}_T^C \le \lim_{\substack{k=n \\ n\to\infty}} P^C(d_L) = \lim_{k\to\infty} \left(\frac{\kappa}{k_1}\right)^k = 0.$$

The numerical result is presented in Table I. It shows that the location privacy information leakage can be minimized with increasing k for a given n. Furthermore, with a proper setting on (k, n), DAPP can defend against collusion attacks even if there is a relatively high ratio of malicious delivery nodes.

F. Data Integrity

In reality, delivery nodes may drop, tamper with, and forge the received data due to their own interests. To deal with these issues, DAPP provides a mechanism to verify the integrity of the received data. The application data server needs k untampered data pieces to recover the original data d and D. In case that at least one data piece in the set of k pieces has been tampered with, we should have $H(d) \neq D$. Therefore, the recovered data can detect whether the data set contains any corrupted data pieces and detect the integrity of the recovered data.

VI. PERFORMANCE EVALUATION AND SIMULATION RESULTS

In urban sensing networks, QoS can be measured by communication delay and delivery ratio. Here, we analyze the communication delay and error rate of DAPP scheme.

A. Communication Delay

The DAPP scheme includes two phases for data forwarding: data forwarding from the source node to the delivery nodes, and from the delivery nodes to the application data server.

In the first phase, let the communication delay for the ith data piece distribution be τ_i, which includes two parts. One part of the delay is introduced to encounter the ith delivery node, which is denoted as γ_i. Theorem 2 proves that the data distribution process is a Poisson process. Hence, γ_i should follow Γ distribution. The probability distribution function for γ_i is given by

$$f_{\gamma_i}(t) = \lambda e^{-\lambda t} \cdot \frac{(\lambda t)^{(n-1)}}{(n-1)!}, \tag{26}$$

TABLE II
ERROR RATE FOR THE REPORTED DATA. $n = 30$

$P_E(\%)$	$k=5$	$k=10$	$k=15$	$k=20$	$k=25$
$p_e=1\%$	2.64×10^{-46}	1.31×10^{-33}	1.27×10^{-22}	4.59×10^{-13}	4.83×10^{-5}
$p_e=2\%$	1.70×10^{-38}	2.52×10^{-27}	7.31×10^{-18}	7.87×10^{-10}	2.51×10^{-3}
$p_e=3\%$	6.20×10^{-34}	1.15×10^{-23}	4.19×10^{-15}	5.70×10^{-8}	2.33×10^{-2}
$p_e=4\%$	1.05×10^{-30}	4.43×10^{-21}	3.65×10^{-13}	1.13×10^{-6}	1.06×10^{-1}
$p_e=5\%$	3.35×10^{-28}	4.39×10^{-19}	1.13×10^{-11}	1.10×10^{-5}	3.28×10^{-1}

where λ is the arrival rate of the Poisson process. The second part is introduced by the inserted constant interval t_α. For the ith delivery node, the delay is equal to it_α.

In the second phase, the delay is the time duration from the time that the delivery node receives a data piece until it reaches the next AP. Suppose the delivery node receives one data piece at time $\hat{t}$ and T is the average time duration that the delivery node travels from one AP to the next AP. The time $\hat{t}$ should be uniformly distributed in the range of T. We denote the delay in the second forwarding phase as t_i for each data piece d_i. The total communication delay $\mathcal{T}_d$ can be computed as follows:

$$\mathcal{T}_d = \min_{1\to k}\{\gamma_i + t_i + it_\alpha \,|i = 1,\ldots,k\}, \tag{27}$$

where $\gamma_i + t_i + it_\alpha$ is the delay for the application data server to receive the ith data component, and $\min_{1\to k}$ is the function to find the kth minimum value of a given set. Equation (27) shows that the overall delay is determined by k when it is large. Table III gives numerical results for various k values.

B. Delivery Ratio

Due to the nature of the wireless communications, the message received may contain errors. Unreliable delivery nodes may even drop the received data pieces for their own interests. To deal with this problem, the proposed scheme introduces redundancy to minimize the error rate for the reporting data. The underlying secret sharing scheme ensures the reported data can be recovered by receiving k or more valid data pieces. The extra $n-k$ data pieces add redundancy that can be employed for data recovery under erroneous scenarios. Let p_e be the error rate or packet loss rate of a single data piece. The overall error rate P_E for the reported data can be computed by the following:

$$P_E = \sum_{i=n-k+1}^{n} \binom{n}{i} p_e^i (1-p_e)^{n-i}. \tag{28}$$

Based on (28), for a given n, the overall error rate P_E decreases when k reduces, and the overall error rate P_E can be minimized by increasing the number of redundant packets. Table II provides numerical results of the overall error rate for various k's.

C. Tradeoff Design and Numerical Simulation Results

In this paper, DAPP is designed to achieve tradeoffs among several conflicting design issues in urban sensing networks. It leverages the relationship between configurable scheme parameters and system performance to provide flexible options for system users. Based on (14) and (27), we have the following equations to determine joint information leakage and communication delay:

TABLE III
AVERAGE COMMUNICATION DELAY AND JOINT INFORMATION LEAKAGE FOR VARIOUS k, WHILE $n = 30, \lambda = 100$ AND $t_\alpha = 20$

$\lambda=1$ $T_\alpha=20$	Location Leakage P_T^D (%)	Communication Delay (s)
$k=5$	2.37	648.7
$k=10$	2.29×10^{-1}	1201.6
$k=15$	3.11×10^{-2}	1749.4
$k=20$	5.20×10^{-3}	2300.7
$k=25$	9.67×10^{-4}	2852.3
$k=30$	2.10×10^{-4}	3418.2

$$\begin{cases} \widetilde{P}_T^D = P(t_s > kt_\alpha) \\ \qquad \times\sum_{i=k}^{n}\left[\left(\frac{1}{4}\right)^{\left\lfloor\frac{2it_\alpha v_S}{\ell}\right\rfloor}\binom{n}{i}P_M^i(1-P_M)^{n-i}\right], \\ \mathcal{T}_d \;= \min_{1\to k}\{\gamma_i + t_i + it_\alpha \,|i = 1,\ldots,k\}. \end{cases} \tag{29}$$

In (29), parameter t_α is designed to prevent one single delivery node from collecting multiple data pieces in a sparse network. It can be set according to the density of population. The two equations show the tradeoff between communication delay and security. For a given t_α and n, the location privacy information leakage can be minimized by increasing the value of k. On the other hand, when k increases, the communication delay also increases, as shown in Table III and Table IV.

Furthermore, for the tradeoff between error rate and security, we have the following equations:

$$\begin{cases} \widetilde{P}_T^D = P(t_s > kt_\alpha) \\ \qquad \times\sum_{i=k}^{n}\left[\left(\frac{1}{4}\right)^{\left\lfloor\frac{2it_\alpha v_S}{\ell}\right\rfloor}\binom{n}{i}P_M^i(1-P_M)^{n-i}\right], \\ P_E = \sum_{i=n-k+1}^{n}\binom{n}{i}p_e^i(1-p_e)^{n-i}. \end{cases} \tag{30}$$

In (30), for a given t_α and n, the error rate can be minimized by decreasing k. However, the direct joint information leakage to the delivery nodes also increases. The numerical results are provided in Table II and Table IV.

In summary, DAPP can achieve tradeoffs among security, communication delay, and delivery ratio. By carefully setting on parameter (n, k, t_α), DAPP can provide an excellent balance between location privacy for participating users and various performance requirements of data users.

D. Computational Complexity

The computational complexity is determined by the overhead in recovery of the secret data using Barycentric Lagrange Interpolation, which is $\mathcal{O}(n)$ for reconstruction of the collected data.

TABLE IV
AVERAGE COMMUNICATION DELAY AND JOINT INFORMATION LEAKAGE FOR VARIOUS (k,n), $t_\alpha = 1$, AND $\lambda = 10$

$Delay\,(s)$ $/P_T^D(\%)$	$k=5$	$k=10$	$k=15$	$k=20$	$k=25$	$k=30$
$n=5$	$219.5/8.89\times10^{-4}$	-	-	-	-	-
$n=10$	$145.2/1.45\times10^{-1}$	$284.8/5.28\times10^{-9}$	-	-	-	-
$n=15$	$140.3/1.13$	$213.8/9.85\times10^{-6}$	$341.3/4.45\times10^{-14}$	-	-	-
$n=20$	$139.8/3.84$	$206.9/3.77\times10^{-4}$	$271.8/4.21\times10^{-10}$	$396.6/3.76\times10^{-19}$	-	-
$n=25$	$139.6/8.71$	$206.7/4.17\times10^{-3}$	$264.7/5.44\times10^{-8}$	$327.4/1.21\times10^{-14}$	$451.2/3.20\times10^{-24}$	-
$n=30$	$139.2/15.6$	$206.2/2.40\times10^{-2}$	$284.3/1.58\times10^{-6}$	$320.3/4.15\times10^{-12}$	$382.7/2.75\times10^{-19}$	$506.1/2.78\times10^{-29}$

If the delivery node only drops data pieces without manipulating, then the Barycentric Lagrange Interpolation algorithm only needs to be implemented once to recover the collected data. However, if the application data server receives modified or corrupted data pieces, it may need to implement the Barycentric Lagrange Interpolation algorithm up to $\binom{n}{k}$ times in the worst case to recover the original data.

Fortunately, the parameters (k,n) with relatively small values are able to decrease the joint information leakage close to zero, as shown in Table IV. As a consequence, it enables the application data server to rebuild the original data in a short time interval. In particular, for $n=25$ and $k=20$, the computer with 2-GHz processor needs at most 0.053s to reconstruct the original data.

E. Numerical Simulation Results

In numerical simulations, we assume v_i follows the uniform distribution in the range of $[V_{\min}, V_{\max}]$. The results of other distributions are similar. To emulate the reality, we set up both dense network and sparse network to show the joint information leakage to the delivery node. In the two scenarios, we select various parameters (k,n,t_α) to show the relationship between the settings and the security performance against intermediate delivery node under side information attack.

In the first scenario, we set the average number of people in the communication range of the source node as $\lambda = 10$, and $t_\alpha = 1$. With proper setting on (k,n), P_M is small enough to make the joint information leakage approximate to zero. The joint information leakage to the delivery node is presented in Table IV.

In the second scenario, we consider a sparse network in the urban area. We set $\lambda = 100$. Through (16) and (17), we can derive that the average number of people in the communication range of source node to be around one. As a consequence, there may be only one neighbor node around the source node to distribute each data piece. In such a scenario, the joint information leakage is determined by $P(t_s \geq kt_\alpha)\cdot(1/4)^{\lfloor 2kt_\alpha v_0/L\rfloor}$. To minimize this value, we raise t_α to 20 and let k increases from 5 to 30. The joint information leakage to the delivery node is shown in Table II and Table IV. The simulation results of the tables can be summarized as follows.

- Location privacy can be preserved in the proposed DAPP scheme under side information attack.
- With proper parameter settings, the joint information leakage can be minimized to approximately zero for each delivery node.
- The joint information leakage, average communication delay, and error rate can be adjusted by selection of system parameters.

VII. CONCLUSION

In this paper, we have proposed a DAPP data-forwarding scheme for people-centric urban sensing networks. It is developed based on Sharmir's secret sharing, dynamic pseudonyms, and a two-phase data-forwarding method. The two-phase forwarding method detaches the connection between the source node and the application data server. Shamir's secret sharing prevents delivery nodes from discovering privacy information in the reported data. The proposed dynamic pseudonym scheme can defend against identity-based side information attack. Both theoretical analysis and simulation results demonstrate that the proposed DAPP can provide an excellent design tradeoff among security, communication delay, and delivery ratio.

APPENDIX
PROOF OF THEOREM 2

In the Manhattan street model, the node arrival and departure at an endpoint of a street segment is a Poisson process. It is straightforward to derive that the arrival process at any points in the street segment parallelizing the endpoint is also a Poisson process. Suppose that the participant A encounters n participants when it stays at the point (x_i, y_i). It stays at the point for a time duration $\delta\,(\delta \to 0)$. The node arrival rate at this point is λ. Therefore, the probability that A meets n participants is

$$p\left(N(\delta+t)-N(t)=n\right) = e^{-\lambda\delta}\cdot\frac{(\lambda\delta)^n}{n!}, \tag{31}$$

where $N(t)$ is the number of participants that node A has encountered from time 0 to time t.

Suppose A needs T seconds to move from (x_i, y_i) to (x_j, y_j). Therefore, the process for A to encounter other participants for time duration T is the sum of encounter processes from location (x_i, y_i) to (x_j, y_j). These meeting processes are mutually independent. Since the number of the independent Poisson processes is T/δ, the sum of multiple Poisson

processes is still a Poisson process. Hence, the new arrival rate of the Poisson process can be computed by the following:

$$\lambda_{\text{sum}} = \sum_{i=1}^{T/\delta} \lambda_i = \frac{T}{\delta} \cdot \lambda. \tag{32}$$

The probability that A meets n participants in time duration T can be calculated as

$$\begin{aligned} p\left(N(t+T) - N(t) = n\right) &= e^{-\lambda_{\text{sum}} \cdot \delta} \cdot \frac{(\lambda_{\text{sum}} \cdot \delta)^n}{n!} \\ &= e^{-\lambda \cdot T} \cdot \frac{(\lambda \cdot T)^n}{n!}. \end{aligned}$$

Therefore, it is a Poisson process. The data distribution is actually equal to the encountering Poisson process. The probability to forward one data piece in a given time duration t_d is

$$p\left(N(t+t_d) - N(t) = 1\right) = e^{-\lambda t_d}(\lambda t_d). \tag{33}$$

However, the source node will insert an interval between two data distributions in the first forwarding phase. Suppose the data piece d_i is forwarded to the delivery node at time t, and the source node insert an interval t_α before forwarding of d_{i+1}. Since the Poisson process is a memoryless process, the number of arrivals occurring in any bounded interval of time after time t is independent of the number of arrivals occurring before time t. Therefore, the probability to forward d_{i+1} in a given time duration t_d is

$$p\left(N(t+t_d+t_\alpha) - N(t+t_\alpha) = 1\right) = e^{-\lambda t_d}(\lambda t_d). \tag{34}$$

Therefore, the data distribution process in the proposed scheme is a Poisson process. ■

REFERENCES

[1] J. Burke *et al.*, "Participatory sensing," in *Proc. 1st Workshop World-Sensor-Web*, Oct. 2006, pp. 1–5.

[2] A. T. Campbell, S. B. Eisenman, N. D. Lane, E. Miluzzo, and R. A. Peterson, "People-centric urban sensing," in *Proc. 2nd Annu. Int. Workshop Wireless Internet*, 2006, pp. 18–31.

[3] R. Du *et al.*, "Effective urban traffic monitoring by vehicular sensor networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 1, pp. 273–286, Jan. 2014.

[4] U. Lee, E. Magistretti, M. Gerla, P. Bellavista, and A. Corradi, "Dissemination and harvesting of urban data using vehicular sensing platforms," *IEEE Trans. Veh. Technol.*, vol. 58, no. 2, pp. 882–901, Feb. 2009.

[5] A. Kapadia, N. Triandopoulos, C. Cornelius, D. Peebles, and D. Kotz, "Anonysense: Opportunistic and privacy-preserving context collection," *Pervasive Comput.*, vol. 5013, pp. 280–297, 2008.

[6] K. L. Huang, S. Kanhere, and W. Hu, "Towards privacy-sensitive participatory sensing," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, Mar. 2009, pp. 1–6.

[7] J. Shi, R. Zhang, Y. Liu, and Y. Zhang, "Prisense: Privacy-preserving data aggregation in people-centric urban sensing systems," in *Proc. IEEE INFOCOM*, Mar. 2010, pp. 758–766.

[8] S. Gao, J. Ma, W. Shi, G. Zhan, and C. Sun, "TRPF: A trajectory privacy-preserving framework for participatory sensing," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 6, pp. 874–887, Jun. 2013.

[9] B. Palanisamy and L. Liu, "Mobimix: Protecting location privacy with mix-zones over road networks," in *Proc. IEEE 27th ICDE*, Apr. 2011, pp. 494–505.

[10] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proc. ACM 1st Int. Conf. Mobile Syst., Appl. Services*, 2003, pp. 31–42.

[11] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Locality-sensitive hashing scheme based on p-stable distributions," in *Proc. 20th Annu. Symp. Comput. Geometry*, 2004, pp. 253–262.

[12] T. Xu and Y. Cai, "Exploring historical location data for anonymity preservation in location-based service," in *Proc. IEEE INFOCOM*, Mar. 2008, pp. 547–555.

[13] C. Ardagna, M. Cremonini, S. De Capitani di Vimercati, and P. Samarati, "An obfuscation-based approach for protecting location privacy," *IEEE Trans. Dependable Secure Comput.*, vol. 8, no. 1, pp. 13–27, Jan. 2011.

[14] K. Vu, R. Zheng, and J. Gao, "Efficient algorithms for k-anonymous location privacy in participatory sensing," in *Proc. INFOCOM*, Mar. 2012, pp. 2399–2407.

[15] S. Gao, J. Ma, W. Shi, and G. Zhan, "Towards location and trajectory privacy protection in participatory sensing," *Mobile Comput., Appl. Services*, vol. 95, pp. 381–386, 2011.

[16] H. Lu, C. S. Jensen, and M. L. Yiu, "Pad: Privacy-area aware, dummy-based location privacy in mobile services," in *Proc. 7th ACM Int. Workshop Data Eng. Wireless Mobile Access*, 2008, pp. 16–23.

[17] H. Kido, Y. Yanagisawa, and T. Satoh, "An anonymous communication technique using dummies for location-based services," in *Proc. 9th Int. Symp. PETS*, 2005, pp. 88–97.

[18] M. Terrovitis and N. Mamoulis, "Privacy preservation in the publication of trajectories," in *Proc. 9th Int. Conf. MDM*, Apr. 2008, pp. 65–72.

[19] E. De Cristofaro and R. Di Pietro, "Adversaries and countermeasures in privacy-enhanced urban sensing systems," *IEEE Syst. J.*, vol. 7, no. 2, pp. 311–322, Jun. 2013.

[20] D. Christin, J. Guillemet, A. Reinhardt, M. Hollick, and S. Kanhere, "Privacy-preserving collaborative path hiding for participatory sensing applications," in *Proc. IEEE 8th Int. Conf. MASS*, Oct. 2011, pp. 341–350.

[21] L. Hu and C. Shahabi, "Privacy assurance in mobile sensing networks: Go beyond trusted servers," in *Proc. IEEE 8th Int. Conf. Pervasive Comput. Commun. Workshops*, Mar. 2010, pp. 613–619.

[22] J. Kim and S. Bohacek, "A survey-based mobility model of people for simulation of urban mesh networks," in *Proc. MeshNets*, 2005, pp. 1– 11.

[23] X. Y. Li and T. Jung, "Search me if you can: Privacy-preserving location query service," in *Proc. INFOCOM*, Apr. 2013, pp. 2760–2768.

[24] V. Vukadinovic, O. R. Helgason, and G. Karlsson, "An analytical model for pedestrian content distribution in a grid of streets," *Math. Comput. Modell.*, vol. 57, no. 11/12, pp. 2933–2944, Jun. 2013.

[25] A. Shamir, "How to share a secret," *Commun. ACM*, vol. 22, no. 11, pp. 612–613, Nov. 1979.

[26] F. Bai, N. Sadagopan, and A. Helmy, "Important: A framework to systematically analyze the impact of mobility on performance of routing protocols for adhoc networks," in *Proc. IEEE INFOCOM*, Mar. 2003, vol. 2, pp. 825–835.

Di Tang received the B.E. degree from Xidian University, Xi'an, China, in 2005. He is currently working toward the Ph.D. degree in electrical and computer engineering with Michigan State University, East Lansing, MI, USA.

His research interests include wireless sensor networks and network security.

Jian Ren (SM'15) received the B.S. and M.S. degrees in mathematics from Shaanxi Normal University, Xi'an, China, and the Ph.D. degree in electrical engineering from Xidian University, Xi'an.

He is currently an Associate Professor with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, USA. His current research interests include cryptography, cloud computing security, network security, energy-efficient sensor network security protocol design, privacy-preserving communications, network coding, distributed network storage, and cognitive networks.

Dr. Ren received the U.S. National Science Foundation Faculty Early Career Development Award in 2009.