Privacy Management and Optimal Pricing in People-Centric Sensing

Mohammad Abu Alsheikh, *Student Member, IEEE,* Dusit Niyato, *Fellow, IEEE,* Derek Leong,
Ping Wang, *Senior Member, IEEE,* and Zhu Han, *Fellow, IEEE*

Abstract—With the emerging sensing technologies such as mobile crowdsensing and Internet of Things (IoT), people-centric data can be efficiently collected and used for analytics and optimization purposes. This data is typically required to develop and render people-centric services. In this paper, we address the privacy implication, optimal pricing, and bundling of people-centric services. We first define the inverse correlation between the service quality and privacy level from data analytics perspectives. We then present the profit maximization models of selling standalone, complementary, and substitute services. Specifically, the closed-form solutions of the optimal privacy level and subscription fee are derived to maximize the gross profit of service providers. For interrelated people-centric services, we show that cooperation by service bundling of complementary services is profitable compared to the separate sales but detrimental for substitutes. We also show that the market value of a service bundle is correlated with the degree of contingency between the interrelated services. Finally, we incorporate the profit sharing models from game theory for dividing the bundling profit among the cooperative service providers.

Index Terms—Data privacy, service pricing, people-centric sensing, mobile crowdsensing, participatory sensing.

I. INTRODUCTION

People-centric sensing incorporates mobile crowdsensing and the Internet of Things (IoT) to provide a platform for people to share ideas, surrounding events, and other sensing data. The collected data is required in creating and updating people-centric services[1] offered to customers over the Internet. However, people-centric data comes with privacy threats which impede crowdsensing participants from providing their true data. A recent survey [1] revealed that 90% of people are concerned about their data privacy and 33% are "unsure" whether to login with their actual identities in IoT devices. Therefore, privacy-awareness and pricing models are required in people-centric services to attain the maximum profit for service providers by jointly optimizing the privacy level and subscription fee.

People-centric services can be sold separately or together as a service bundle. Specifically, people-centric services can be interrelated with a degree of contingency. There is a joint demand for *complementary services* as both services are jointly required by the customers, e.g., sentiment analysis and activity tracking. On the other hand, *substitute services* are comparable in their functionality, e.g., sentiment analysis using two data analytics algorithms. To attract customers to buy the service bundle, the bundle is offered at a lower subscription fee compared to the sum of subscription fees of the separately sold services. In this paper, we examine three important questions that arise regarding pricing and bundling of people-centric services. Firstly, what are the optimal subscription fee and privacy level required for profit maximization? Secondly, is service bundling effective for profit maximization in people-centric services and does it produce more gross profit compared to the separate sales? Thirdly, how is the bundling profit divided among the interrelated services?

This paper provides a framework for optimal pricing and privacy management in people-centric services. We assume that service providers are rational, refuse to offer their services unless they can recover the total data cost, and seek to maximize their own gross profits from selling people-centric services. The key contributions of this paper can be summarized as follows:

- We first present a model to define the utility of data in people-centric sensing. Using real-world datasets, we show that this utility model captures the inverse correlation between the privacy level and service quality where data analytics is extensively utilized.
- We formulate a pricing scheme and profit maximization model for separately selling privacy-aware people-centric services. The data is collected from crowdsensing participants and used in training and updating people-centric services which are offered to customers for a subscription fee. The optimal subscription fee and privacy level are optimized to maximize the gross profit of service providers.
- We then propose a bundling scheme for virtually packaging privacy-aware people-centric services as complements or substitutes. The subscription fee of the service bundle and the privacy levels of the two services are optimized by a profit maximization model. We introduce a profit allocation model for sharing the profit resulting from the service bundle among the individual bundled services. The efficient and fair profit allocations encourage collaboration among service providers in providing service bundles of complementary services. Nonetheless,

M. Abu Alsheikh, D. Niyato, and P. Wang are with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (emails: mohammad027@e.ntu.edu.sg, dniyato@ntu.edu.sg, and wangping@ntu.edu.sg). D. Leong is with the Institute for Infocomm Research, Singapore 138632 (email: dleong@i2r.a-star.edu.sg). Z. Han is with the School of Electrical and Computer Engineering, University of Houston, TX, USA 77004 (email: zhan2@uh.edu).

[1]Examples of people-centric services include Waze (https://www.waze.com) for online traffic monitoring and PatientsLikeMe (https://www.patientslikeme.com) for healthcare experience sharing on treatments and symptoms.

bundling of substitute services is detrimental and yields low profit allocations.

The rest of this paper is structured as follows. Section II reviews the related work. Then, the system model and a brief overview of people-centric sensing are presented in Section III. Section IV introduces the optimization model of separately selling privacy-aware people-centric services. Next, the profit maximization models of selling bundled complementary and substitute services with privacy awareness are presented in Section V. Section VI presents and analyzes the experimental results. The paper is finally concluded in Section VII.

II. RELATED WORK

People are the primary focus in people-centric sensing, which has applications in transportation systems [2], assistive healthcare [3], and urban monitoring [4], just to name a few. In this section, we first review related work on pricing models in sensing and communication systems. Then, we discuss the crucial issue of privacy awareness in people-centric sensing. Finally, we review related work on pricing and incentive mechanisms for mobile crowdsensing.

A. Pricing in Sensing and Communication Systems

Pricing models ensure financial stability and resiliency in sensing and communication systems. The authors in [5] presented a cooperative pricing model for Internet providers offering the Internet service under one coalition. The cooperative pricing increases the profit and encourages the Internet providers to upgrade their network connections. A pricing scheme based on the customer data usage of Internet services was introduced in [6]. Unlike flat-rate pricing, the usage-based pricing enables a fair allocation of the Internet resource among the customers. The authors in [7] presented a pricing model of accessing femtocell and macrocell by mobile devices which enables high service quality and maximizes the profit of network operators. The authors in [8] proposed a pricing and transmission scheduling models to maximize the profit of accessing a wireless network by mobile customers. The customer demand is modeled as a Markov chain where applying only two price options is found sufficient for each demand state.

The pricing models of people-centric services is more challenging compared to other communication systems. Specifically, the resources and utility of people-centric services are not easily measured as other systems, e.g., the bandwidth and connection speed are easily defined for an Internet service.

B. Privacy-Awareness in People-Centric Sensing

People-centric sensing incorporates the IoT and mobile crowdsensing and considers the privacy of people. In [3], the authors pointed out that the data privacy is a key challenge for people-centric sensing in assistive healthcare. The data traces in healthcare systems typically include personal habits and traits of users and introduce the risk of privacy disclosure. The authors in [4] discussed people-centric sensing as a data collection method in urban areas. The crowdsensing participants apply additive and non-additive aggregation models, such as averaging, to ensure that their true data cannot be disclosed. In [9], anonymous data collection in people-centric sensing was presented as a privacy preserving model. The privacy model is applied with low resource demand of CPU and network bandwidth as mobile devices are resource-constrained. Additionally, there are other works related to location privacy in people-centric sensing, e.g., [10]–[12], which prevent localization attacks and protect the privacy of participants in spatial tasks. Spatial cloaking, adding noise, and rounding are commonly used in the literature for location obfuscation.

There are a few models to define the data privacy including k-anonymity [13], l-diversity [14], and differential privacy [15]. The k-anonymity model requires that a person cannot be identified from at least a group of $k-1$ other people. However, k-anonymity does not guarantee privacy against background knowledge attacks, e.g., merging quasi-sensitive attributes with other datasets. The l-diversity model [14] addresses background knowledge attacks and ensures that any quasi-sensitive group has at least l "well-represented" values. Differential privacy [15] does not include assumptions on the adversary's background knowledge. Instead, differential privacy requires that the probability distribution of shared data from a privacy preserving method does not change by adding one person's data. This ensures that the adversary cannot recognize a particular person from differentially-private data. Random noise is generally added to the raw data to meet the definition of differential privacy [16]–[18].

C. Pricing and Incentive Mechanisms for Mobile Crowdsensing

Pricing and incentive mechanisms are required to encourage participation in data collection. A reward-based incentive mechanism for mobile crowdsensing was presented in [19]. The reservation wages of the participants are utilized to reduce the total data cost by selecting the sufficient set of participants with the lowest rates. In [20], the participants are paid according to their reliability. The reliability is defined as a probabilistic process and measured based on the historical records of the participants in completing crowdsensing tasks. The authors in [21] considered the heterogeneity of crowdsensing participants and proposed asymmetric payment model which encourages competition among the participants. In [22], the authors introduced a profit maximization and pricing model to optimize the amount of data that should be bought from the sensing participants.

None of the existing papers on people-centric sensing in the literature consider the problem of jointly optimizing the pricing and privacy level in people-centric services where data analytics is heavily applied. Moreover, existing works do not consider bundling interrelated people-centric services as complements or substitutes. Therefore, there is a practical demand for privacy-aware pricing, bundling, and profit allocation models which are the major contributions of this paper.

III. People-Centric Big Data: System Model

In this section, we first discuss the system model of people-centric sensing and services considered in this paper. We then briefly present the market strategy of bundling people-centric services. Finally, we introduce the tradeoff between the privacy level and service quality in data analytics perspectives.

A. People-Centric Services

Figure 1 shows the system model of people-centric sensing and services. People share their crowdsensing data through a massive system of mobile devices and IoT gadgets. In particular, the proliferation of sensors, e.g., cameras, microphones, and accelerometers, in mobile devices enables people to participate in cooperative sensing of events and phenomena. Besides, people intelligence can be incorporated in the sensing process which helps in collecting complex and rich data. Other data regarding people can also come from conventional sensor networks. The network connections in people-centric sensing include various technologies such as cellular and Wi-Fi networks. In the following we describe the major entities in people-centric services under consideration:

- *Crowdsensing participants* are the source of the data. We consider a people-centric service with N crowdsensing participants where each participant i produces privacy preserving data denoted as follows:

$$y_i = x_i + z_i, \quad i = 1, \ldots, N \tag{1}$$

where y_i is the noisy data, x_i is the true data, and z_i is the added noise component. We assume that the noise components $\{z_i\}_{i=1}^{N}$ are independent Gaussian random components with zero mean and a variance of σ_z^2. We also denote this data in the vector form as $\mathbf{y} \in \mathbb{R}^N$, $\mathbf{x} \in \mathbb{R}^N$, and $\mathbf{z} \in \mathbb{R}^N$ such that $\mathbf{z} \sim \mathbb{N}\left(0, \sigma_z^2 \mathbb{I}_N\right)$ where $\mathbb{I}_N$ is the identity matrix of size N. To attain high service quality, the participant stochastically sends its true data $\mathbf{x}$ to the people-centric service with a probability of $\mathbb{P}(\text{true data})$. For the remainder of this paper, we define the privacy level r to be equal to the probability of sending the noisy data $\mathbf{y}$ instead of the true data $\mathbf{x}$ such that $\mathbb{P}(\text{true data}) = 1 - r$. Our system model is general and can incorporate any privacy definition such as k-anonymity [13], l-diversity [14], and differential privacy [15]. Each participant, whether it is a member of the public or data warehouses, has a *reservation wage* c which is the lowest payment required to work as a data collector.

- A *service provider* buys people-centric data from the crowdsensing participants and applies data analytics to build the people-centric service. This service is hosted at one of the cloud computing platforms such as Microsoft Azure and Amazon Web Services (AWS). To cover the operation cost, the service provider charges a "subscription fee" p_s to customers who access the people-centric service. Moreover, the service provider decides the "privacy level" r at which the data should be collected by the crowdsensing participants. For gross profit maximization,

the service provider jointly optimizes the subscription fee and privacy level of its service.

- *Customers* are the users of the people-centric service. Each customer has a *reservation price* θ which is the maximum price at which that particular customer will buy the people-centric service. A customer considers both the reservation price θ, service quality u, and subscription fee p_s when making its buying decision. In particular, a customer buys the service if the inequality $\theta \geq \frac{p_s}{u}$ holds.

As summarized in Table I, this people-centric sensing model overcomes the limitations of conventional sensing systems based on sensor networks only. However, people-centric sensing comes with the privacy challenge which should be considered in optimal pricing and profit maximization.

B. Bundling Interrelated Services

People-centric services can be interrelated and sold as one service bundle[2]. Figure 2 illustrates the interrelation among people-centric services from the perspectives of customers. Complementary services are associated and concurrently required to achieve the objectives of customers. For example, both sentiment analysis [23] and activity tracking [24] are typically required to provide in-depth understanding of human-intense mobile systems. On the other hand, substitute services have similar or comparable functionalities which decrease the customer's willingness in buying both services. If a customer buys one of the substitute services, that customer will probably not buy its paired service. For example, sentiment analysis using the data analytics models of deep learning [25] and random forests [26] are substitute. In some scenarios, the customers buy both substitute services to improve the performance of service quality, e.g., a mixture of models in ensemble learning [27].

C. The Quality-Privacy Tradeoff

There are several reasons for examining the privacy and optimal pricing of people-centric services. Firstly, privacy is a common concern of people. Secondly, reservation wages of crowdsensing participants are correlated with the utility of data and service quality. The quality of data analytics is inversely proportional to privacy level [28]. Thirdly, the customers infer both the service quality and subscription fee when deciding whether to buy a people-centric service. The utility function of data $u(\cdot)$ in people-centric services should meet the following empirical assumptions:

- $u(\cdot)$ is nonnegative. This is rational as the service quality cannot be negative.
- $u(\cdot)$ is inversely proportional to the privacy level $r \in [0, 1]$ such that $\frac{\partial u(\cdot)}{\partial r} < 0$. This empirical assumption is required as increasing the privacy level decreases the quality of data analytics [28].
- $u(\cdot)$ is convex and decreases at an increasing rate over the privacy level such that $\frac{\partial^2 u(\cdot)}{\partial r^2} < 0$. This assumption

---

[2]Product bundling is an effective marketing strategy of selling products in one package, e.g., Microsoft Office includes Microsoft Word, Excel, and PowerPoint.
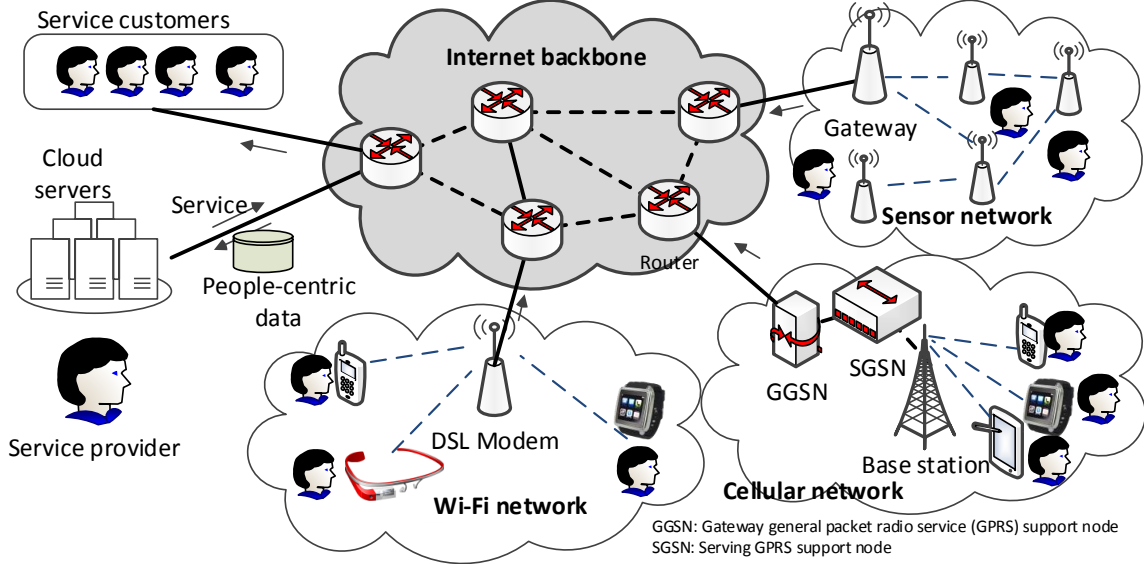
Fig. 1: System model of people-centric data collection and service development.

TABLE I: Comparison of people-centric sensing and conventional sensing.

| ASPECT	PEOPLE-CENTRIC SENSING	CONVENTIONAL SENSING
Deployment	Mobile devices owned by participants	Sensor nodes typically owned by service providers
People engagement	Human in the loop	Machines only
Mobility	Move with people	Static or limited mobility
Key challenge	Data privacy	Energy conservation

(a) Complementary services (complements)

(b) Substitute services (substitutes)
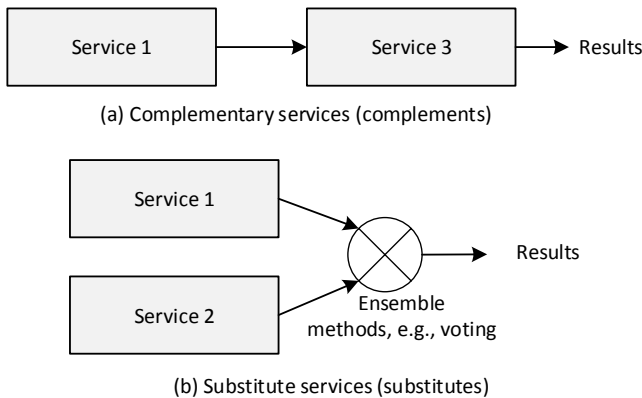
Fig. 2: Interrelated services as complements and substitutes.

reflects the empirical change of service quality at varying privacy levels.

Based on these empirical assumptions and to facilitate our optimization modeling, we propose the following utility function:

$$u(r; \alpha) = \alpha_1 - \alpha_2 \exp(\alpha_3 r), \qquad (2)$$

where r is the privacy level. α_1, α_2, and α_3 are the curve fitting parameters of the utility function to real-world experiments, i.e., the ground truth. Big data platforms, e.g., Apache Mahout [29] and MLlib [30], can be used for running the real-world experiments at scale. In particular, a set of B real-world experiments $\left\{\left(r^{(i)}, \tau^{(i)}\right)\right\}_{i=1}^{B}$ are executed at varying privacy level $r^{(i)}$ resulting in the real-world service quality of $\tau^{(i)}$, where $r^{(i+1)} > r^{(i)} \geq 0$. α_1, α_2, and α_3 are obtained by minimizing the residuals of a nonlinear least squares fitting as follows:

$$\underset{\alpha}{\text{minimize}} \ \sum_{i=1}^{B} \left\| u(r^{(i)}; \alpha) - \tau^{(i)} \right\|^2. \qquad (3)$$

(3) can be solved iteratively to find the best fitting parameters α [31]. We denote $u(r; \alpha)$ as u whenever it does not cause confusion. In Section VI-B, we show the validity of (2) in capturing the quality-privacy tradeoff of people-centric services trained on real-world datasets.

Figure 3 shows the key components of the optimal pricing and privacy management framework proposed in this paper. These components are executed iteratively. The framework is initiated by defining the data utility using the form expressed in (2). Then, the profit maximization models are executed to obtain the optimal subscription fee and privacy level. These profit maximization models are presented in Sections IV and V for separate and bundling sales, respectively. For bundling, the profit allocation models presented in Section V-D are performed. Then, the service provider decides whether the service bundling is effective to attain the maximum profit. The key notations and definitions used throughout the paper are defined in Table II.
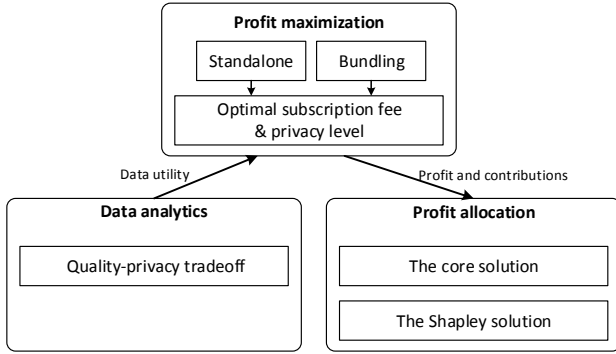
IV. Optimal Pricing in People-Centric Services

In this section, we first present the market model of selling people-centric service separately. Then, we introduce the profit maximization model with privacy awareness. Finally, the closed-form solutions of the subscription fee and privacy level are derived and proved to be globally optimal.

A. Gross Profit Maximization

The system model under consideration in this section is shown in Figure 1. The service provider collects data from crowdsensing participants at a privacy level r. The data is essential to train and update a people-centric service offered to customers paying a subscription fee p_s. The gross profit $F(\cdot)$ of selling the people-centric service can be defined mathematically as follows:

$$F(r, p_s) = \underbrace{M p_s \mathbb{P}\left(\theta \geq \frac{p_s}{u}\right)}_{\text{Subscription revenue}} - \underbrace{N c \mathbb{P}\left(\text{true data}\right)}_{\text{Total data cost}}, \quad (4)$$

where M is the number of potential customers, p_s is the service subscription fee, r is the privacy level, c is the reservation wage of crowdsensing participants, and N is the number of potential crowdsensing participants. The gross profit $F(\cdot)$ is the difference between the subscription revenue and total data cost. The operational cost of the service, such as the computing cost, is neglected. The first term of (4) defines the subscription revenue resulting from offering the service at a subscription fee of p_s and service quality u. $\mathbb{P}\left(\theta \geq \frac{p_s}{u}\right) = 1 - \Phi_\theta\left(\frac{p_s}{u}\right)$ is the probability for a customer to buy the service after inferring both p_s and u which can be calculated from the complementary cumulative distribution function. The total data cost is defined in the second term of (4) to be proportional to the probability of sending the true data $\mathbb{P}\left(\text{true data}\right)$. This is rational as the service quality and gross profit are negatively affected by increasing the privacy level, and thus the service provider should also pay less for the noisy data. Assuming that θ follows a uniform distribution over the interval $[0, 1]$, (4) can be written as follows:

$$F(r, p_s) = M p_s \left(1 - \frac{p_s}{\alpha_1 - \alpha_2 \exp\left(\alpha_3 r\right)}\right) - N c \left(1 - r\right). \quad (5)$$

The profit maximization problem can be formulated as follows:

$$\begin{aligned}
\underset{r, p_s}{\text{maximize}} \quad & F(r, p_s) \\
\text{subject to} \quad & C1: p_s \geq 0, \\
& C2: r \geq 0.
\end{aligned} \quad (6)$$

The objective of (6) is to maximize the gross profit by jointly optimizing p_s and r. The constraints $C1$ and $C2$ are required to ensure nonnegative solutions of p_s and r, respectively. We next provide the closed-form solution (p_s^*, r^*) of this profit maximization problem and prove their global optimality.

B. Optimal Subscription Fee and Privacy Level

We apply the Karush-Kuhn-Tucker (KKT) conditions which are sufficient for primal-dual optimality of concave func-

Fig. 3: Components of the optimal pricing and privacy management framework for people-centric sensing.

TABLE II: List of key symbols and notations.

Notation	Definition
p_s	Subscription fee required to access a people-centric service under separate selling
r	Privacy level at which the data is collected from crowdsensing participants
$F(r, p_s)$	Gross profit resulting from the separate sales of a people-centric service under p_s and r
$u(r; \alpha)$	Service quality with curve fitting parameters α and privacy level r. u_k is the service quality of the k-th service in a service bundle
M	Number of customers willing to buy a people-centric service
N	Number of crowdsensing participants
c	Reservation wage of one crowdsensing participant
θ	Reservation price defining the maximum price at which a customer is willing to buy a people-centric service. θ has a cumulative distribution function of Φ_θ
$\mathcal{L}(\cdot)$	Lagrangian of the profit function $F(\cdot)$
D_k	k-th leading principal minor of the profit function's Hessian matrix
γ	Degree of contingency of two interrelated services
S_b	Service bundle
p_b	Subscription fee of a service bundle S_b containing two complementary or substitute services
$G_c(r_1, r_2, p_b)$	Gross bundling profit of complementary services with privacy levels r_1 and r_2 offered at a subscription fee of p_b
$G_s(r_1, r_2, p_b)$	Gross bundling profit of substitute services with privacy levels r_1 and r_2 offered at a subscription fee of p_b
φ_k	Payoff allocation assigned using the core solution to the k-th service in S_b
η_k	Payoff allocation assigned using the Shapley value concept to the k-th service in S_b

tions [32]. Based on (6), we formulate the Lagrangian dual function as follows:

$$\mathcal{L}\left(r, p_s, \lambda_1, \lambda_2\right) = F(r, p_s) + \lambda_1 p_s + \lambda_2 r, \qquad (7)$$

where $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ are the Lagrange multipliers (dual variables) associated with the constraints C_1 and C_2, respectively.

Proposition 1. *The closed-form solutions p_s^* and r^* of (6) can be derived as follows:*

$$p_s^* = \frac{M\alpha_1\alpha_3 - 4Nc}{2M\alpha_3}, \qquad (8)$$

$$r^* = \frac{1}{\alpha_3}\log\left(\frac{4Nc}{M\alpha_2\alpha_3}\right), \qquad (9)$$

where $\lambda_1 = \lambda_2 = 0$.

Proof: To achieve this result, the first derivatives of (7) with respect to p_s and r are found as follows:

$$\frac{\partial\mathcal{L}\left(\cdot\right)}{\partial p_s} = M + \lambda_1 - \frac{2Mp_s}{\alpha_1 - \alpha_2\exp\left(\alpha_3 r\right)}, \qquad (10)$$

$$\frac{\partial\mathcal{L}\left(\cdot\right)}{\partial r} = Nc + \lambda_2 - \frac{M\alpha_2\alpha_3 p_s^2\exp\left(\alpha_3 r\right)}{\left(\alpha_1 - \alpha_2\exp\left(\alpha_3 r\right)\right)^2}. \qquad (11)$$

The closed-form solutions in (8) and (9) with inactive inequality constraints ($\lambda_1 = \lambda_2 = 0$) can then be deduced by setting both derivatives to zero and solving the resulting system of equations. ∎

Proposition 2. *$F(r, p_s)$ is concave. The closed-form solutions p_s^* and r^* given in (8) and (9), respectively, are globally optimal.*

Proof: We use Sylvester's criterion as a sufficient condition to show that the Hessian matrix of $F(r, p_s)$ is negative semidefinite, and hence the concavity of $F(r, p_s)$ can be deduced [32]. In particular, the Hessian matrix $\mathbf{H}_F$ of $F(r, p_s)$ is found as in (12), shown at the top of the next page. Let D_k be the k-th leading principal minor of $\mathbf{H}_F$, where $k = 1, 2$. $\mathbf{H}_F$ is negative semidefinite if $(-1)^k D_k \geq 0$ [32]. The leading principal minors of $\mathbf{H}_F$ are

$$D_1 = -\frac{2M}{\alpha_1 - \alpha_2\exp\left(\alpha_3 r\right)} \leq 0, \qquad (13)$$

$$D_2 = \frac{2M^2\alpha_2\alpha_3^2 p_s^2\exp\left(\alpha_3 r\right)}{\left(\alpha_1 - \alpha_2\exp\left(\alpha_3 r\right)\right)^3} \geq 0, \qquad (14)$$

which alternate in sign with D_1 being nonpositive. Therefore, $\mathbf{H}_F$ is negative semidefinite and $F(r, p_s)$ is concave. By concavity, the closed-form solutions expressed in (8) and (9) are globally optimal. ∎

C. Special Case: Fixed Privacy Level

We next discuss the special case when the service provider cannot control the privacy level, e.g., r is fixed by a legislative court[3]. In such a case, the profit of the service provider can

[3]The European Commission (http://ec.europa.eu), for example, regularly revises a set of regulations to protect the data privacy of citizens in the European Union.

be defined as in (5) with r being fixed. For profit maximization, the service provider responds by selecting the optimal subscription fee as deduced in the following proposition.

Proposition 3. *When the privacy level r is fixed by an external entity, the optimal subscription fee is found as follows:*

$$p_s^* = \frac{\alpha_1 - \alpha_2\exp\left(\alpha_3 r\right)}{2}, \qquad (15)$$

which is globally optimal.

Proof: The second derivatives of $F(r, p_s)$ given in (5) with respect to p_s is defined as follows:

$$\frac{\partial^2 F\left(\cdot\right)}{\partial p_s^2} = -\frac{2M}{\alpha_1 - \alpha_2\exp\left(\alpha_3 r\right)} \leq 0, \qquad (16)$$

which is always nonpositive. Thus, $F(r, p_s)$ is concave and the solution in (15) of the fixed privacy problem is globally optimal. ∎

V. INTERRELATED PEOPLE-CENTRIC SERVICES

People-centric services can be interrelated as complements and substitutes as shown in Figure 2. The joint optimization of the subscription fee and privacy levels in a service bundle is introduced in this section. Firstly, we present the system model and define the degree of contingency in service bundling. Secondly, we present the profit maximization models and the closed-form solutions of selling service bundles as complements and substitutes, respectively. The closed-form solutions are also shown to be globally optimal. Finally, we define the profit shares that should be allocated to each service within the bundle.

A. Market Model and Degree of Contingency

We consider the marketing strategy of virtually bundling two services denoted as service S_1 and service S_2 into a bundle denoted as service bundle S_b as shown in Figure 4. We identify the degree of contingency between the two services as γ. γ indicates the customer interest in obtaining the two services S_1 and S_2 as a service bundle S_b. We incorporate the definition of γ from microeconomics as follows [33]:

$$\gamma = \frac{\theta_b - (\theta_1 + \theta_2)}{\theta_1 + \theta_2}, \qquad (17)$$

where θ_b, θ_1, and θ_2 are the reservation prices of the service bundle S_b, the standalone service S_1, and the standalone service S_2, respectively. The service bundle S_b can be classified into two types as follows:

- $\theta_b \geq (\theta_1 + \theta_2)$ and hence $\gamma \geq 0$: When θ_b is greater than or equal to the summation of θ_1 and θ_2, S_1 and S_2 are complementary services. For example, the customers are willing to buy both services S_1 and S_2 as each service has a unique functionality.
- $\theta_b < (\theta_1 + \theta_2)$ and hence $\gamma < 0$: When θ_b is less than the summation of θ_1 and θ_2, S_1 and S_2 are substitute services. For example, the customers are not willing to buy both services S_1 and S_2 as they are similar and comparable in functionality.

$$\mathbf{H}_F = \begin{bmatrix} -\dfrac{2M}{\alpha_1 - \alpha_2 \exp(\alpha_3 r)} & -\dfrac{2M\alpha_2\alpha_3 p_s \exp(\alpha_3 r)}{(\alpha_1 - \alpha_2 \exp(\alpha_3 r))^2} \\ -\dfrac{2M\alpha_2\alpha_3 p_s \exp(\alpha_3 r)}{(\alpha_1 - \alpha_2 \exp(\alpha_3 r))^2} & -\dfrac{2M\alpha_2^2\alpha_3^2 p_s^2 \exp(2\alpha_3 r)}{(\alpha_1 - \alpha_2 \exp(\alpha_3 r))^3} - \dfrac{M\alpha_2\alpha_3^2 p_s^2 \exp(\alpha_3 r)}{(\alpha_1 - \alpha_2 \exp(\alpha_3 r))^2} \end{bmatrix}. \tag{12}$$
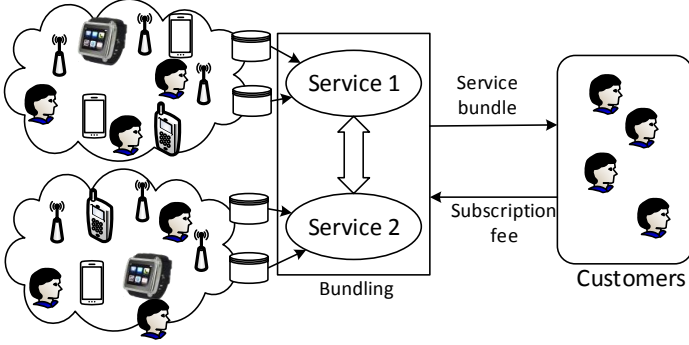

Fig. 4: System model of bundled people-centric services.

The customer demand on buying S_b can be found as in Figure 5. u_k is the service quality of the k-th service in S_b. The shaded regions represent the probability of buying S_b. p_b is the subscription fee of S_b. Each point in the figure represents the reservation price pair (θ_1, θ_2) of the two services in the bundle. For both complementary and substitute services, a customer will buy S_b if (θ_1, θ_2) lies above the decision line $(1+\gamma)(\theta_1 u_1 + \theta_2 u_2) = p_b$. Moreover, the customers of substitute services will also buy S_b when (θ_1, θ_2) lies at the right side of the line $\frac{p_b}{u_1}$ and above the line $\frac{p_b}{u_2}$. We observe that the customers are more willing to buy S_b containing complementary services than substitute services.

B. Complementary People-Centric Services ($\gamma \geq 0$)

S_1 and S_2 are complements when the reservation price of the service bundle S_b is greater than the total reservation price of the standalone services $\theta_b \geq (\theta_1 + \theta_2)$. The gross profit of S_b containing complementary services can be defined as follows:

$$G_c(r_1, r_2, p_b) = \underbrace{Mp_b \mathbb{P}\left((1+\gamma)(\theta_1 u_1 + \theta_2 u_2) > p_b\right)}_{\text{Total subscription revenue}}$$
$$\underbrace{- N\left(c_1(1-r_1) + c_2(1-r_2)\right)}_{\text{Total data cost of } S_1 \text{and } S_2}, \tag{18}$$

where M is the number of potential customers willing to buy the service bundle S_b. p_b is the subscription fee of S_b. r_1 and r_2 are the privacy levels of S_1 and S_2, respectively. N is the number of crowdsensing participants. c_1 and c_2 are the reservation wages of the crowdsensing participants in S_1 and S_2, respectively. $u_1 = \alpha_1 - \alpha_2 \exp(\alpha_3 r_1)$ is the service quality of S_1 and $u_2 = \beta_1 - \beta_2 \exp(\beta_3 r_2)$ is the service quality of S_2 as formulated in (2). $\mathbb{P}\left((1+\gamma)(\theta_1 u_1 + \theta_2 u_2) > p_b\right)$ is the probability of buying S_b as defined by the shaded area of complementary services in Figure 5. Assuming that θ_1 and

θ_2 follow a uniform distribution, (18) can be re-written as follows:

$$G_c(r_1, r_2, p_b) = Mp_b \left(1 - \frac{0.5 p_b^2}{(1+\gamma)^2 u_1 u_2}\right) - Nc_1(1-r_1) - Nc_2(1-r_2). \tag{19}$$

The profit maximization problem by selling two complements in S_b is then expressed as follows:

$$\begin{aligned} \underset{r_1, r_2, p_b}{\text{maximize}} \quad & G_c(r_1, r_2, p_b) \\ \text{subject to} \quad & C3 : p_b \geq 0, \\ & C4 : r_1 \geq 0, \\ & C5 : r_2 \geq 0, \end{aligned} \tag{20}$$

where $C3$, $C4$, and $C5$ are the optimization constraints required to ensure nonnegative solutions of p_b, r_1, and r_2, respectively. The Lagrangian dual function of (20) is derived as:

$$\mathcal{L}_c(r_1, r_2, p_b, \lambda, \lambda_4, \lambda_5) = G_c(r_1, r_2, p_b) + \lambda_3 p_b + \lambda_4 r_1 + \lambda_5 r_2, \tag{21}$$

where λ_3, λ_4, and λ_5 are the Lagrange multipliers.

Proposition 4. *The closed-form solutions p_b^*, r_1^*, and r_2^* of (20) are given as follows:*

$$p_b^* = \frac{0.5 A_1}{M \alpha_3 \beta_3}, \tag{22}$$

$$r_1^* = \frac{1}{\alpha_3} \log\left(\frac{13.5 N^2 c_1 c_2}{M^2 \alpha_2 \alpha_3 \beta_1 \beta_3 (\gamma^2 + 2\gamma + 1)} + \frac{2.25 N c_1 A_1}{M^2 \alpha_2 \alpha_3^2 \beta_1 \beta_3 (\gamma^2 + 2\gamma + 1)}\right), \tag{23}$$

$$r_2^* = \frac{1}{\beta_3} \log\left(\frac{13.5 N^2 c_1 c_2}{M^2 \alpha_1 \alpha_3 \beta_2 \beta_3 (\gamma^2 + 2\gamma + 1)} + \frac{2.25 N c_2 A_1}{M^2 \alpha_1 \alpha_3 \beta_2 \beta_3^2 (\gamma^2 + 2\gamma + 1)}\right), \tag{24}$$

where

$$\begin{aligned} A_1 = \Big[& \frac{8}{2} \alpha_1 \beta_1 M^2 \alpha_3^2 \gamma^2 \beta_3^2 + \frac{16}{3} \alpha_1 \beta_1 M^2 \alpha_3^2 \gamma \beta_3^2 \\ & + \frac{8}{2} \alpha_1 \beta_1 M^2 \alpha_3^2 \beta_3^2 + 9 N^2 \alpha_3^2 c_2^2 - 18 N^2 \alpha_3 c_1 c_2 \beta_3 \\ & + 9 N^2 c_1^2 \beta_3^2 \Big]^{0.5} - 3 N \alpha_3 c_2 - 3 N \beta_3 c_1. \end{aligned} \tag{25}$$

Proof: This result can be proved by taking the first derivatives of (21) with respect to r_1, r_2, and p_b. From the three resulting equations and with inactive constraints ($\lambda_1 = \lambda_2 = \lambda_3 = 0$), the closed form solutions can be obtained. ∎
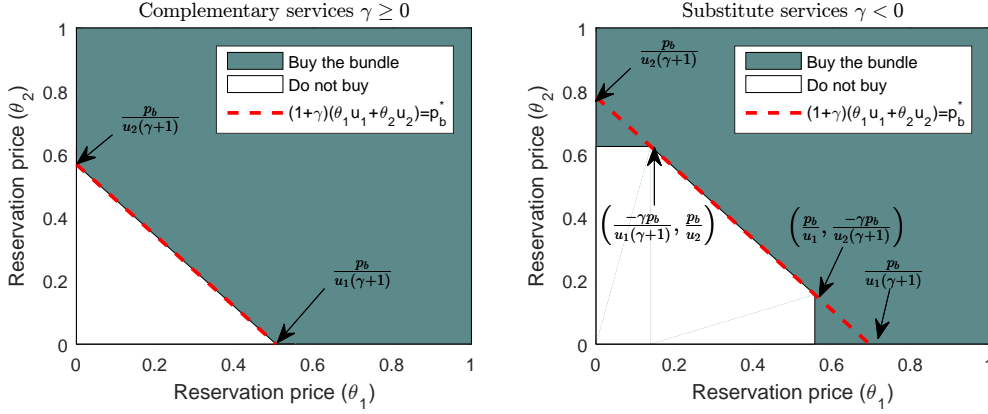
Fig. 5: Customer demand on the service bundles of complements and substitutes.

Proposition 5. *The profit function $G_c(r_1, r_2, p_b)$ defined in (19) for complementary people-centric services is concave. The closed-form solutions p_b^*, r_1^*, and r_2^* given in (22), (23), and (24), respectively, are globally optimal.*

Proof: The Hessian matrix $\mathbf{H}_G$ of $G_c(r_1, r_2, p_b)$ is found as in (26), shown at the top of the next page. The leading principal minors can then be obtained as in (27)-(29), shown at the top of the next page, where

$$A_2 = M^3 \alpha_1 \alpha_2 \alpha_3^2 p_b^7 \beta_2^2 \beta_3^2 exp\left(\alpha_3 r_1\right) exp\left(2\beta_3 r_2\right)$$
$$+ M^3 \alpha_2^2 \alpha_3^2 p_b^7 \beta_1 \beta_2 \beta_3^2 exp\left(\beta_3 r_2\right) exp\left(2\alpha_3 r_1\right)$$
$$- 2M^3 \alpha_1 \alpha_2 \alpha_3^2 p_b^7 \beta_1 \beta_2 \beta_3^2 exp\left(\alpha_3 r_1\right) exp\left(\beta_3 r_2\right). \quad (30)$$

$G_c(r_1, r_2, p_b)$ is concave as $D_1 \leq 0$, $D_2 \geq 0$, and $D_3 \leq 0$ when $A_2 \leq 0$. Accordingly, the closed-form solutions are globally optimal. ■

We next present the optimal solutions to the profit maximization problem with fixed privacy levels for S_1 and S_2.

1) Fixed Privacy Level: When the privacy levels are enforced by an external legislation entity, we have the following proposition.

Proposition 6. *When the privacy levels r_1 and r_2 are fixed by an external entity, the optimal subscription fee of the service bundle is expressed as follows:*

$$p_b^* = \frac{0.82\left(\gamma + 1\right)\left(\alpha_1 - \alpha_2 exp\left(\alpha_3 r_1\right)\right)\left(\beta_1 - \beta_2 exp\left(\beta_3 r_2\right)\right)}{\left(\left(\alpha_1 - \alpha_2 exp\left(\alpha_3 r_1\right)\right)\left(\beta_1 - \beta_2 exp\left(\beta_3 r_2\right)\right)\right)^{0.5}}. \quad (31)$$

(31) is globally optimal.

Proof: The second derivatives of $G_c(r_1, r_2, p_b)$ defined in (19) with respect to p_b is derived as follows:

$$\frac{\partial^2 G_c\left(\cdot\right)}{\partial p_b^2} =$$
$$\frac{-3Mp_b}{\left(\alpha_1 - \alpha_2 exp\left(\alpha_3 r_1\right)\right)\left(\beta_1 - \beta_2 exp\left(r_2 \beta_3\right)\right)\left(\gamma + 1\right)^2} \leq 0, \quad (32)$$

which is nonpositive. Thus, the profit maximization problem with a fixed privacy level is concave and the solution in (31) is globally optimal. ■

C. Substitute People-Centric Service $(\gamma < 0)$

As they have comparable functionality, substitute services are only required to obtain better overall service quality, e.g., predictive performance. Services S_1 and S_2 are called substitutes when the reservation price of the service bundle S_b less than the total reservation price of the separate sales $\theta_b < (\theta_1 + \theta_2)$. Bundling substitute services yields in the following gross profit:

$$G_c(r_1, r_2, p_b) = Mp_b \mathbb{P}\left(\begin{array}{c}[(1+\gamma)(\theta_1 u_1 + \theta_2 u_2) > p_b] \\ \cup \left[\theta_1 \geq \frac{p_b}{u_1}\right] \cup \left[\theta_2 \geq \frac{p_b}{u_2}\right]\end{array}\right)}_{\text{Total subscription revenue}}$$
$$- \underbrace{Nc_1\left(1 - r_1\right) - Nc_2\left(1 - r_2\right)}_{\text{Total data cost of } S_1 \text{and } S_2}. \quad (33)$$

$\mathbb{P}\left([(1+\gamma)(\theta_1 u_1 + \theta_2 u_2) > p_b] \cup \left[\theta_1 \geq \frac{p_b}{u_1}\right] \cup \left[\theta_2 \geq \frac{p_b}{u_2}\right]\right)$ is the probability for a customer to buy S_b which can be defined by the shaded area of substitute services in Figure 5. Then, (33) can be re-written as follows:

$$G_s(r_1, r_2, p_b) = Mp_b\left(1 - \frac{0.5p_b^2 + \gamma^2 p_b^2}{\left(1 + \gamma\right)^2 u_1 u_2}\right)$$
$$- Nc_1\left(1 - r_1\right) - Nc_2\left(1 - r_2\right). \quad (34)$$

The profit maximization problem of selling two substitute services in S_b is expressed as follows:

$$\begin{aligned}
\underset{r_1, r_2, p_b}{\text{maximize}} \quad & G_s(r_1, r_2, p_b) \\
\text{subject to} \quad & C6 : p_b \geq 0, \\
& C7 : r_1 \geq 0, \\
& C8 : r_2 \geq 0.
\end{aligned} \quad (35)$$

The objective is maximizing the gross profit of S_b under the constraints $C6$, $C7$, and $C8$ for nonnegative solutions in p_b^*, r_1^*, and r_2^*, respectively.

Proposition 7. *The profit function $G_s(r_1, r_2, p_b)$ defined in (34) for substitute people-centric services is concave. The closed-form solutions p_b^*, r_1^*, and r_2^* are given in (36), (37), and (37), respectively,*

$$\mathbf{H}_G = \begin{bmatrix} -\frac{0.5M\alpha_2\alpha_3^2 p_b^3 exp(\alpha_3 r_1)u_1}{(\gamma+1)^2 u_1^3 u_2} & -\frac{0.5M\alpha_2\alpha_3 p_b^3 \beta_2 \beta_3 exp(\alpha_3 r_1) exp(r_2\beta_3)}{(\gamma+1)^2 u_1^2 u_2^2} & -\frac{1.5M\alpha_2\alpha_3 p_b^2 exp(\alpha_3 r_1)}{(\gamma+1)^2 u_1^2 u_2} \\ -\frac{0.5M\alpha_2\alpha_3 p_b^3 \beta_2 \beta_3 exp(\alpha_3 r_1) exp(r_2\beta_3)}{(\gamma+1)^2 u_1^2 u_2^2} & -\frac{0.5M p_b^3 \beta_2 \beta_3^2 exp(r_2\beta_3)u_2}{(\gamma+1)^2 u_1 u_2^3} & -\frac{1.5M p_b^2 \beta_2 \beta_3 exp(r_2\beta_3)}{(\gamma+1)^2 u_1 u_2^2} \\ -\frac{1.5M\alpha_2\alpha_3 p_b^2 exp(\alpha_3 r_1)}{(\gamma+1)^2 u_1^2 u_2} & -\frac{1.5M p_b^2 \beta_2 \beta_3 exp(r_2\beta_3)}{(\gamma+1)^2 u_1 u_2^2} & -\frac{3M p_b}{(\gamma+1)^2 u_1 u_2} \end{bmatrix}. \tag{26}$$

$$D_1 = -\frac{0.5M\alpha_2\alpha_3^2 p_b^3 exp\left(\alpha_3 r_1\right)\left(\alpha_1 + \alpha_2 exp\left(\alpha_3 r_1\right)\right)}{\left(\gamma+1\right)^2 \left(\beta_1 - \beta_2 exp\left(\beta_3 r_2\right)\right)\left(\alpha_1 - \alpha_2 exp\left(\alpha_3 r_1\right)\right)^3}, \tag{27}$$

$$D_2 = \frac{0.25M^2\alpha_2\alpha_3^2 p_b^6 \beta_2 \beta_3^2 exp\left(\alpha_3 r_1 + r_2\beta_3\right)\left(\alpha_1\beta_1 + \alpha_2\beta_1 exp\left(\alpha_3 r_1\right) + \alpha_1\beta_2 exp\left(r_2\beta_3\right)\right)}{\left(\gamma+1\right)^4 \left(\beta_1 - \beta_2 exp\left(r_2\beta_3\right)\right)^4 \left(\alpha_1 - \alpha_2 exp\left(\alpha_3 r_1\right)\right)^4}, \tag{28}$$

$$D_3 = \frac{0.375A_2}{\left(\gamma+1\right)^6 \left(\beta_1 - \beta_2 exp\left(\beta_3 r_2\right)\right)^5 \left(\alpha_1 - \alpha_2 exp\left(\alpha_3 r_1\right)\right)^5}. \tag{29}$$

$$p_b^* = -\frac{0.5A_3}{M\alpha_3\beta_3}, \tag{36}$$

$$r_1^* = \frac{1}{\alpha_3} \log\left(\frac{13.5\left(c_1 c_2 N^2\gamma^2 + c_1 c_2 N^2\right)}{M^2\alpha_2\alpha_3\beta_1\beta_3\left(\gamma^2 + 2\gamma + 1\right)} - \frac{2.25\left(Nc_1\gamma^2 + Nc_1\right)A_3}{M^2\alpha_2\alpha_3^2\beta_1\beta_3\left(\gamma^2 + 2\gamma + 1\right)}\right), \tag{37}$$

$$r_2^* = \frac{1}{\beta_3} \log\left(\frac{13.5Nc_2\left(Nc_1\gamma^2 + Nc_1\right)}{M^2\alpha_1\alpha_3\beta_2\beta_3\left(\gamma^2 + 2\gamma + 1\right)} - \frac{2.25Nc_2\left(\gamma^2 + 1\right)\left(A_3\right)}{M^2\alpha_1\alpha_3\beta_2\beta_3^2\left(\gamma^2 + 2\gamma + 1\right)}\right). \tag{38}$$

where

$$A_3 = 3N\alpha_3 c_2 + 3Nc_1\beta_3 - 3N\alpha_3 c_2 - 3Nc_1\beta_3$$
$$+ \frac{1}{9\left(\gamma^2 + 1\right)^2}\big[8\alpha_1\beta_1 M^2\alpha_3^2\gamma^2\beta_3^2 + 16\alpha_1\beta_1 M^2\alpha_3^2\gamma\beta_3^2$$
$$+ 8\alpha_1\beta_1 M^2\alpha_3^2\beta_3^2 + 27N^2\alpha_3^2 c_2^2\gamma^2 + 27N^2\alpha_3^2 c_2^2$$
$$- 54N^2\alpha_3 c_1 c_2\gamma^2\beta_3 - 54N^2\alpha_3 c_1 c_2\beta_3 + 27N^2 c_1^2\gamma^2\beta_3^2$$
$$+ 27N^2 c_1^2\beta_3^2\big]^{0.5}. \tag{39}$$

These closed-form solutions are globally optimal.

Proof: The proof is similar to the ones of Propositions 4 and 5 and will be omitted due to the space limit. ∎

D. Profit Sharing

A service bundle can be formed by two service providers forming a bundling coalition $\mathcal{K}$. We next present a profit sharing model to divide the bundling profit among the cooperative providers.

1) Core Solution: Let φ_k indicate the profit share of the service provider S_k, where $k \in \mathcal{K}$. The core solution $\mathcal{C}$ is defined as follows [34]:

$$\mathcal{C} = \left\{\varphi \mid \underbrace{\sum_{k\in\mathcal{K}}\varphi_k = G_{\mathcal{K}}^*}_{\text{group rationality}} \text{ and } \underbrace{\sum_{k\in\mathcal{S}}\varphi_k \geq F_{\mathcal{S}}^*, \mathcal{S} \subseteq \mathcal{K}}_{\text{individual rationality}}\right\} \tag{40}$$

where $G_{\mathcal{K}}^*$ is the bundling profit and $F_{\mathcal{S}}^*$ is the profit resulting from selling the services separately.

The core solution $\mathcal{C}$ can contain an infinite number of possible share allocations, can be empty, or lead to unfair share allocations when considering the contributions of services in S_b. We next present the Shapley value concept which provides a fair and single solution to the profit sharing problem of S_b.

2) The Shapley Value Solution: For each service S_k, where $k \in \mathcal{K}$, forming the bundle S_b, the Shapley value solution $\eta = (\eta_1, \eta_2)$ ensures fairness and assigns a payoff η_k defined as [34]:

$$\eta_k = \sum_{\mathcal{S}\subseteq\mathcal{K}\backslash\{k\}} \underbrace{\frac{|\mathcal{S}|!\left(|\mathcal{K}| - |\mathcal{S}| - 1\right)!}{|\mathcal{K}|!}}_{\text{probability of random ordering}} \underbrace{\left(G_{\mathcal{K}}^* - F_{\mathcal{S}}^*\right)}_{\text{marginal contribution}}. \tag{41}$$

The first term defines the random order of joining the bundle. The second term defines the marginal contribution of each service on increasing the bundling profit.

VI. EXPERIMENTAL RESULTS

In this section, we first present three people-centric services which are trained using real-world datasets. We also analyze the quality of the services when deep learning [25] and random forests [26] are utilized as data analytics algorithms. We then introduce numerical results of selling the services separately. Finally, we evaluate the bundling models for selling complementary and substitute services, respectively.

A. People-Centric Services and Bundles

Using real-world datasets, we design the following people-centric services:

- **Service S_1 (sentiment analysis using deep learning)**: Using the Sentiment140 dataset [23], we develop a service to predict people's sentiment from social networking tweets. The sentiment can be either positive or negative. People post tweets which typically include personal information, and privacy awareness is reasonably required. We use $629,145$ tweet samples for model training and $419,431$

tweet samples for model testing and quality calculation. We assume that the reservation wage of each crowdsensing participant is 0.2[4].

- **Service** S_2 *(sentiment analysis using random forests)*: This service is similar to service S_1 above, except in using a random forest classifier instead of deep learning.
- **Service** S_3 *(activity tracking using random forests)*: This service enables the tracking of human activities using the accelerometer sensors of mobile devices. We use the Actitracker dataset [24] containing a time series of $1,098,207$ data points. We divide the accelerometer time series into overlapping window frames resulting in $23,072$ training and $5,768$ testing samples. The predicted activities are walking, jogging, walking upstairs, walking downstairs, sitting, and standing. We set the reservation wage of each crowdsensing participant as 0.1.

These people-centric services can be sold separately or inter-related in service bundles. We consider the following bundling scenarios:

- **Bundle** S_{b1} *(S_1 and S_3)*: Section VI-D considers the economic strategy of virtually packaging services S_1 and S_3 into one service bundle. Services S_1 and S_3 are complementary as both services are typically required to provide in-depth understanding of mobile users. We assume that the degree of contingency is $\gamma = 0.1$ which indicates the high customer willingness in acquiring both services at once.
- **Bundle** S_{b2} *(S_1 and S_2)*: In Section VI-E, we analyze services S_1 and S_2 as substitutes because they have comparable functionality, i.e., both services S_1 and S_2 are used for sentiment analysis, but they differ in the data analytics algorithm. To reflect the low customer willingness of buying comparable services, the degree of contingency is set as $\gamma = -0.1$.

We set the number of crowdsensing participants to $N = 100$ and the number of customers to $M = 1000$.

B. The Quality-Privacy Tradeoff

Figure 6 shows the quality-privacy models of S_1, S_2, and S_3, respectively. We observe three major results. Firstly, the service quality decreases as the privacy level increases. This is clear as increasing the privacy level results in higher data distortion. Secondly, it can be also noted that the real data fits the quality function defined in (2). Thirdly, the service quality of S_1 and S_2 are different even though they use the same dataset. This is due to the different data analytics algorithms used in S_1 and S_2.

C. Standalone Sales

We use service S_1 to evaluate the profit maximization model for selling people-centric services as a standalone product. From Figure 6, the quality-privacy fitting parameters of S_1 are $\alpha_1 = 0.822$, $\alpha_2 = 0.004$, and $\alpha_3 = 2.813$.

---

[4]We use monetary units for all payment, cost, and profit analysis in the experimental results. Actual currency, such as the United States dollar, can be applied without affecting the optimization models or results.

1) Gross Profit Optimization: Figure 7 shows the gross profit $F^*(r, p_s)$ of S_1 defined in (5) under varied privacy level r and subscription fee p_s. When the subscription fee is high, the profit decreases as fewer customers will be willing to pay the subscription fee. When the subscription fee is low, more customers will buy S_1. However, the gross profit falls down due to the low subscription fee. Likewise, a high privacy level results in a low service quality and fewer customers will be accordingly interested in the service of poor quality. A low privacy level results in a high service quality, but gross profit will decrease due to the high spending in buying the true data from the crowdsensing participants. The optimal settings of the subscription fee $p_s^* = 0.406$ and privacy level $r^* = 0.62$ can be found using the closed-form solutions in (8) and (9), respectively. Then using (5), the maximum profit is calculated as $F(r^*, p_s^*) = 195.5$.

2) The Impact of Reservation Wage: In Figure 8, we consider the impact of varying the reservation wage of the crowdsensing participants on the gross profit $F^*(\cdot)$, privacy level r^*, subscription fee p_s^*, and total data cost. Firstly, there is an inverse correlation between the reservation wage and the gross profit. Specifically, when the reservation wage is increased, the total data cost will increase up to $c = 0.15$ and the profit will accordingly decrease. Increasing the reservation wage beyond $c = 0.15$ yields a fall of the total data price as a rational service provider will intensively increase the privacy level as defined in (9). Secondly, we note that the reservation wage and subscription fee are also inversely proportional. In particular, the service provider reduces the subscription fee to attract more customers due to the degradation in the service quality.

3) The Impact of Customer Base: Figure 9 shows the gross profit $F^*(\cdot)$, privacy level r^*, subscription fee p_s^*, and total data cost under varied number of customers. When the number of customers increases, the gross profit and subscription fee increase as the benefit of the increased demand. Moreover, the service provider decreases the privacy level to collect more true data which increases the service quality and total data cost.

4) Fixed Privacy Level: In some scenarios, the service provider does not control the privacy level as discussed in Section IV-C, e.g., due to legislation rules. Instead, the service provider only specifies the subscription fee as in (15) to gain the maximum gross profit. In Figure 10, we analyze the gross profit $F^*(\cdot)$, subscription revenue, subscription fee p_s^*, and total data cost of S_1 at varied privacy level r. We observe two important results. Firstly, the subscription revenue, subscription fee, and total data cost are inversely correlated with the privacy level. This is expected as increasing the privacy level negatively affects the service quality and fewer customers will be interested in buying S_1. Besides, the total data cost will decrease when the privacy level is high. Secondly, we note that the gross profit increases up to $r = 0.62$, then it decreases due to the extreme loss of customers at the high privacy levels $r > 0.62$.

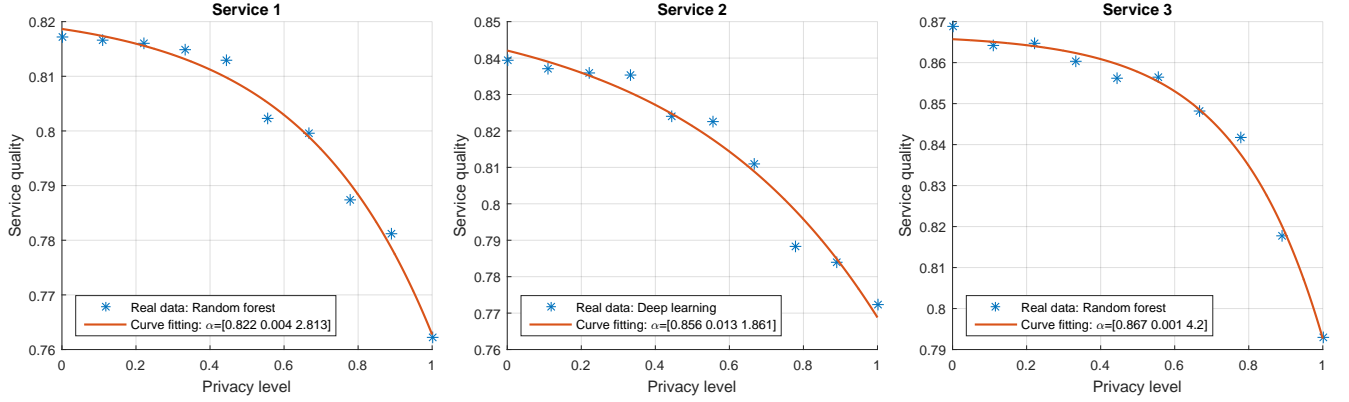Fig. 6: The prediction quality of the services S_1, S_2, and S_3 (from left to right) under varied privacy levels.
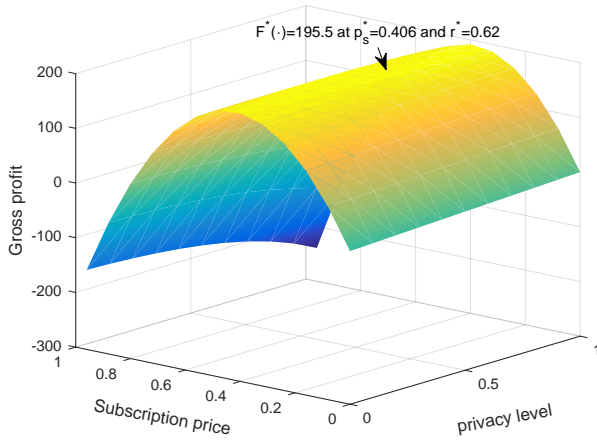

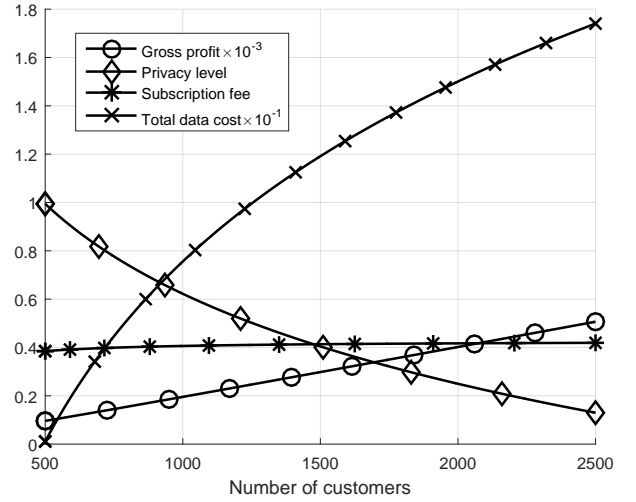Fig. 7: Gross profit of S_1 under varied privacy level r and subscription fee p_s.

Fig. 9: Impacts of the customer base size M on the gross profit $F^*(\cdot)$, privacy level r^*, subscription fee p_s^*, and total data cost.
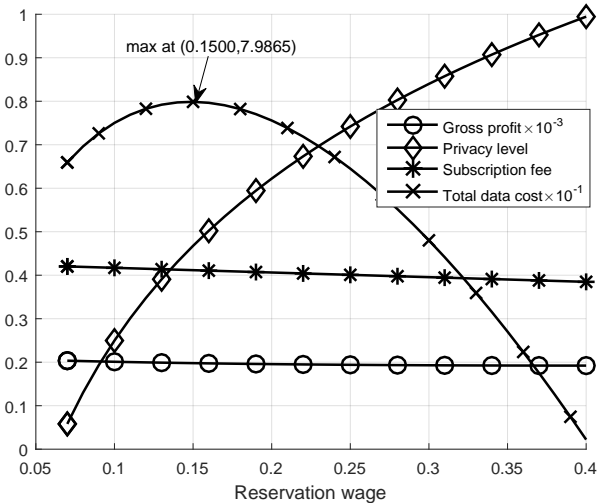

Fig. 8: Impacts of the reservation wage c on the gross profit $F^*(\cdot)$, privacy level r^*, subscription fee p_s^*, and total data cost.

D. Complementary People-Centric Services

We consider bundling S_1 and S_3 as complementary services into S_{b1}. From Figure 6, the fitting parameters of S_1 are $\alpha_1 = 0.822$, $\alpha_2 = 0.004$, and $\alpha_3 = 2.813$. For S_3, the fitting parameters are $\beta_1 = 0.867$, $\beta_2 = 0.001$, and $\beta_3 = 4.2$. We first analyze the bundling profit and the impacts of the different parameters on S_{b1}. We then present the payoff allocations among S_1 and S_3 based on the importance of each service on the sales of S_{b1}.

1) Gross Profit Optimization: The gross bundling profit $G_c(r_1, r_2, p_b)$ defined in (19) is presented in Figure 11. When the subscription fee p_b and the privacy levels r_1 and r_2 are either high or low, the gross profit goes down. Specifically, fewer customers will buy overpriced or poor quality service bundles. Likewise, S_{b1} makes a low profit when the subscription fee and privacy level are low due to the low revenue and high total data cost, respectively. The optimal settings $p_b^* = 0.754$,
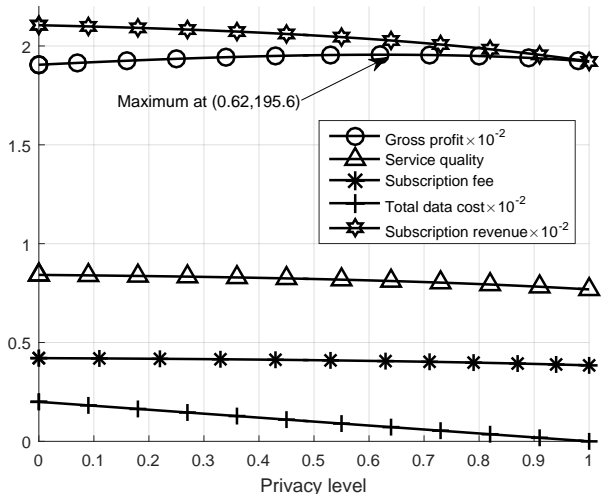
Fig. 10: Gross profit $F^*(\cdot)$, subscription revenue, subscription fee p_s^*, and total data cost under varied privacy level r.

$r_1^* = 0.513$, and $r_2^* = 0.499$ can be obtained using the closed-form solutions given in (22), (23), and (24), respectively. The optimal gross profit of S_{b1} is $G_c(r_1^*, r_2^*, p_b^*) = 487.84$ which is greater than that of selling services S_1 and S_3 as standalone products with $F_1(r^*, p_s^*) = 195.5$ and $F_3(r^*, p_s^*) = 206.02$, respectively. Thus, the rational service providers will decide to build S_{b1} and stop selling S_1 and S_3 as standalone services.

2) Demand Boundaries: Figure 12 shows the demand boundary of S_{b1} in the reservation price spaces. The customers buy S_{b1} when the customer valuations lie above the decision line $(1+\gamma)(\theta_1 u_1 + \theta_2 u_2) = p_b^*$, where $u_1 = 0.82$, $u_2 = 0.859$, $\gamma = 0.1$, and $p_b^* = 0.754$. The customers do not buy S_{b1} when their valuations are below the decision line.

3) The Impact of Contingency Degree: In Figure 13, we analyze the gross profit $G_c^*(\cdot)$, privacy levels r_1^* and r_2^*, subscription fee p_b^*, and total data cost under varied degree of contingency γ. The total data cost of service bundle S_{b1} includes the data costs of S_1 and S_3 as expressed in (19). Firstly, we note that the gross bundling profit is proportional to the degree of contingency. This is clear as the high degree of contingency indicates strong interrelation between S_1 and S_3. Thus, the customers are more interested in buying both services together. Secondly, the subscription fee of S_{b1} is increased to meet any increase in the degree of contingency. The resulting increase in the gross profit motivates the service provider to enhance the overall service quality by decreasing the privacy levels r_1^* and r_2^*.

4) The Impact of Reservation Wage: We consider the impact of varying the reservation wage of service S_1 on the optimal pricing and profits of service bundle S_{b1}. We observe two important results from Figure 14. Firstly, the bundling profit $G_c^*(\cdot)$ goes down when the reservation wage c_1 increases. This is due to the increased data cost of S_1 as defined in (19). Secondly, in order to minimize the total data cost, the privacy level r_1^* of S_1 is increased. The privacy level r_2^* of S_2 is also slightly increased but at a lower rate than

r_1^*. These results can also be deduced from the closed-form solutions of r_1^* and r_2^* in (23) and (24), respectively.

5) Profit Sharing: The bundling profit can be divided between services S_1 and S_3 as shown in Figure 15. The feasible payoffs guarantee that the summation of payoffs does not exceed the bundling profit $\eta_1 + \eta_3 \leq G_c(r_1^*, r_2^*, p_b^*) = 487.84$. The efficient payoffs assign the allocations such that the total payoff is equal to the bundling profit $\eta_1 + \eta_3 = G_c(r_1^*, r_2^*, p_b^*) = 487.84$. The core solution defined in (40) ensures that the payoff allocations of either S_1 or S_3 cannot be improved by leaving the bundle and selling services separately. Finally, the Shapley value solution defined in (41) assigns fair payoff allocations based on the importance of each service forming S_{b1}.

We next study the impacts of varying the reservation wage c_1 of S_1 on the profit shares from S_{b1}. Figure 16 shows the profit resulting from offering S_1 and S_3 separately and jointly as S_{b1}. The profit allocations in S_{b1} is found using the Shapley value solution defined in (41). Two important observations can be made. Firstly, the gross profit falls as c_1 is increased. This has negative effects on the profit of S_3 in both the bundling and separate sales. The maximum-to-minimum profit difference of S_3 in the bundling and separate sales are 13.19 and 11.36, respectively. Secondly, we observe that higher profit allocations can be obtained from S_{b1} for S_1 and S_3 compared to the separate sales. Thus, the providers of services S_1 and S_3 have a monetary incentive in making the service bundle S_{b1} regardless of the data cost. This result shows that the fairness of the Shapley value solution is crucial for stable service bundling in people-centric services.

E. Substitute People-Centric Services

As substitute services, we next consider combining services S_1 and S_2 in into the service bundle S_{b2}. As shown in Figure 6, the fitting parameters of S_1 are $\alpha_1 = 0.822$, $\alpha_2 = 0.004$, and $\alpha_3 = 2.813$. For S_2, the fitting parameters are $\beta_1 = 0.856$, $\beta_2 = 0.013$, and $\beta_3 = 1.861$.

1) Demand Boundaries: Figure 17 presents the demand on S_{b2} consisting of substitute services. There are three decision boundaries. Firstly, the customers buy the bundle if their valuations lie above and to the right of the decision line $(1 + \gamma)(\theta_1 u_1 + \theta_2 u_2) = p_b^*$, where $u_1 = 0.811$, $u_2 = 0.793$, $\gamma = -0.1$, and $p_b^* = 0.58$. Secondly, the customers buy S_{b2} if their valuation θ_1 of S_1 is greater than or equal to $\frac{p_b}{u_1}$. Thirdly, the customers buy S_{b2} if their valuations θ_2 of S_2 are greater than or equal to $\frac{p_b}{u_2}$.

2) The Impact of Contingency Degree: Interrelated products are modeled as substitutes when $\gamma < 0$. Figure 18 shows that when the degree of contingency is decreased, the gross profit $G_s^*(\cdot)$, subscription fee p_b^*, and total data cost decrease. This correlation is expected as decreasing γ indicates high similarity among S_1 and S_2. Thus, the customer valuations of the resulting bundle decrease and the subscription fee moves to lower values accordingly.

3) Profit Sharing: Bundling substitute services is detrimental for the gross profit compared to the separate sales as shown in Figure 19. In particular, the customer valuation of S_{b2}
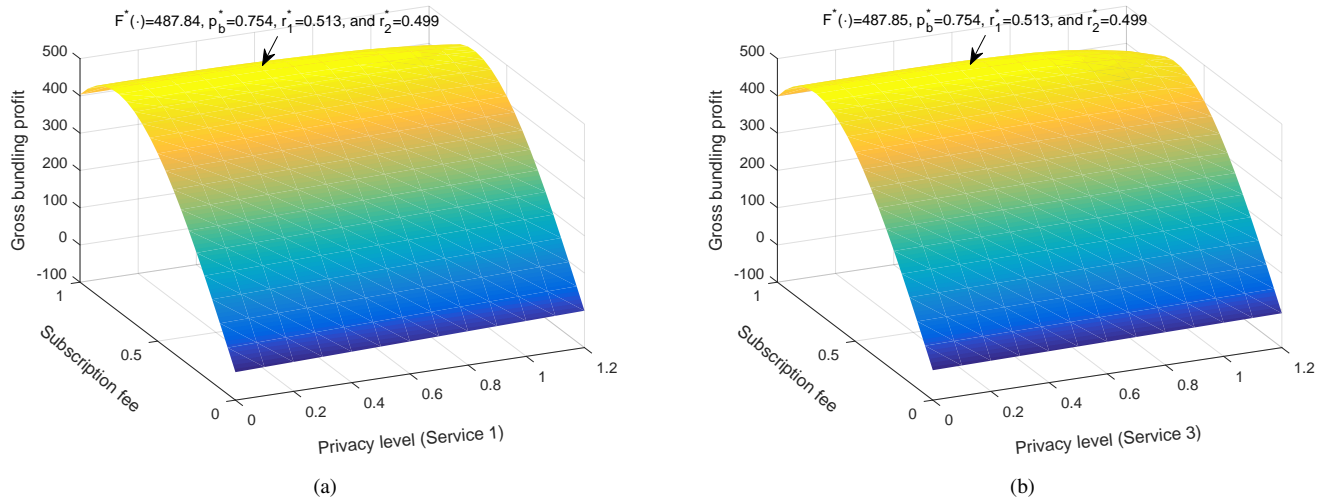
(a)

(b)

Fig. 11: Gross profit $G_c(r_1, r_2, p_b)$ by bundling S_1 and S_3 into the service bundle S_{b1} under varied privacy levels r_1 and r_2 and subscription fee p_b.
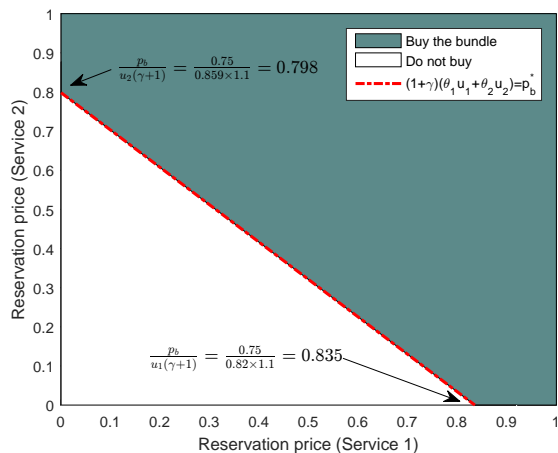

Fig. 12: Demands on S_{b1} containing both S_1 and S_3.

Fig. 13: Impacts of the degree of contingency γ on the gross profit $G_c^*(\cdot)$, privacy levels r_1^* and r_2^*, subscription fee p_b^*, and total data cost.

containing similar and comparable services is reasonably low. The bundling profit, therefore, falls below the total profits under the separate sales of S_1 and S_2 $G_s(r_1^*, r_2^*, p_b^*) < F_1(r^*, p_s^*) + F_2(r^*, p_s^*)$. The rational service providers will decide to sell S_1 and S_2 separately.

VII. CONCLUSION AND FUTURE WORK

In this paper, we have presented the profit maximization and pricing models for selling people-centric services separately and as service bundles. We have firstly modeled the tradeoff between the service quality and privacy level from data analytics perspectives. Specifically, the service quality has been shown to be inversely proportional to the privacy level. For separately offered services, the service provider jointly optimizes the privacy level and subscription fee to maximize the gross profit resulting from offering services to a set of customers. For service bundling, the people-centric services are bundled
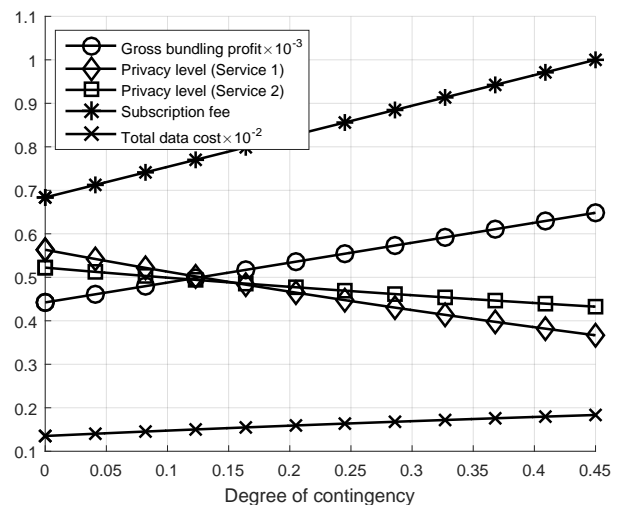
as complementary and substitute services. Accordingly, the privacy levels of the two services and subscription fee are optimized to gain the maximum gross profit for the service provider. Finally, we have presented a model for sharing the resulting bundling profit among the cooperative people-centric services.

For the future work, the heterogeneity of the crowdsensing participants and competitive markets can be included in the profit maximization models.
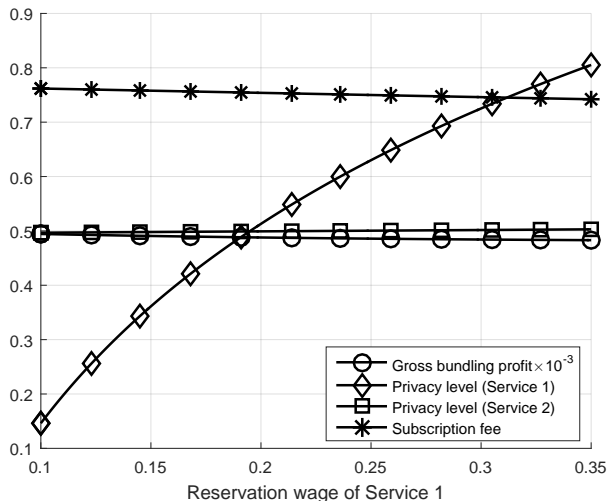
Fig. 14: Impacts of the reservation wage c_1 on the gross profit $G_c^*(\cdot)$, privacy levels r_1^* and r_2^*, and subscription fee p_b^*.
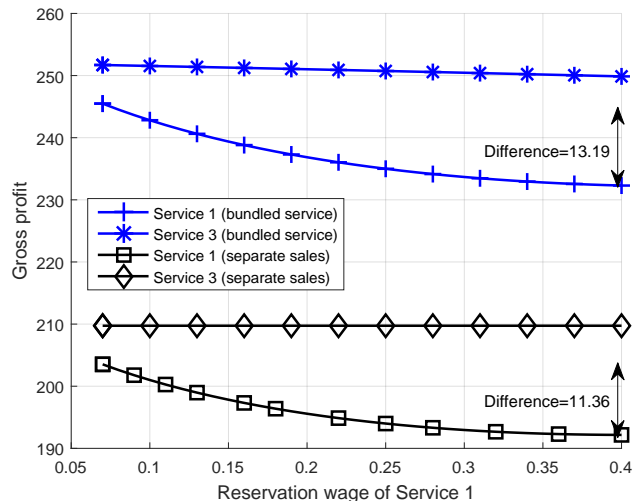

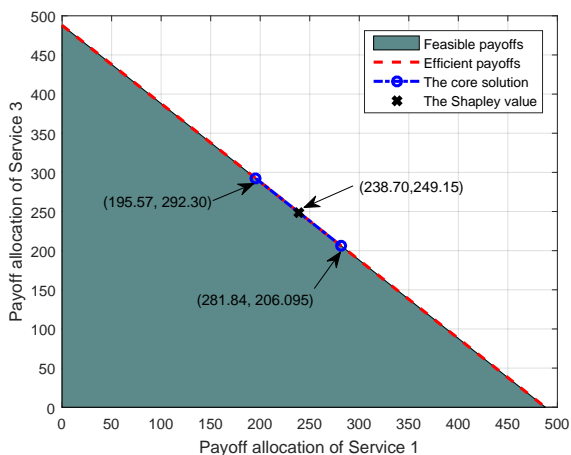Fig. 16: Payoff allocation under varied reservation wage c_1 of S_1.

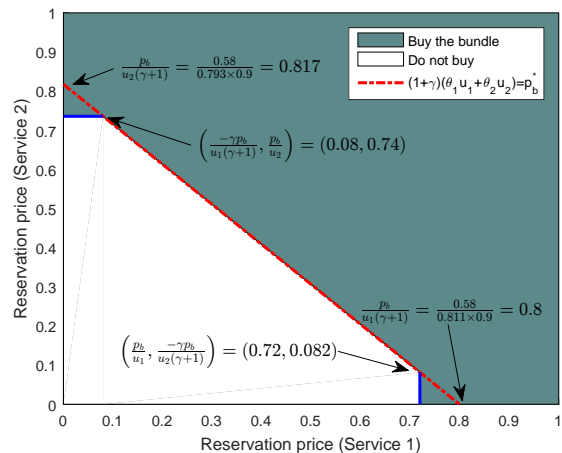Fig. 15: Payoff allocation of the bundling profit from S_{b1} among S_1 and S_3.

Fig. 17: Demand boundaries of S_{b2}.

REFERENCES

[1] Gigya Inc, "The 2015 state of consumer privacy & personalization," *White Paper*, 2015.

[2] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1624–1639, December 2011.

[3] T. Giannetsos, T. Dimitriou, and N. R. Prasad, "People-centric sensing in assistive healthcare: Privacy challenges and directions," *Security and Communication Networks*, vol. 4, no. 11, pp. 1295–1307, February 2011.

[4] R. Zhang, J. Shi, Y. Zhang, and C. Zhang, "Verifiable privacy-preserving aggregation in people-centric urban sensing systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 268–278, September 2013.

[5] L. He and J. Walrand, "Pricing and revenue sharing strategies for Internet service providers," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 5, pp. 942–951, May 2006.

[6] R. T. Ma, "Usage-based pricing and competition in congestible network service markets," *IEEE/ACM Transactions on Networking*, vol. 24, no. 5, pp. 3084–3097, October 2016.

[7] L. Duan, J. Huang, and B. Shou, "Economics of femtocell service provision," *IEEE Transactions on Mobile Computing*, vol. 12, no. 11, pp. 2261–2273, November 2013.

[8] L. Huang and M. J. Neely, "The optimality of two prices: Maximizing revenue in a stochastic communication system," *IEEE/ACM Transactions on Networking*, vol. 18, no. 2, pp. 406–419, April 2010.

[9] C. Cornelius, A. Kapadia, D. Kotz, D. Peebles, M. Shin, and N. Triandopoulos, "Anonysense: Privacy-aware people-centric sensing," in *Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services*. ACM, 2008, pp. 211–224.

[10] J. Krumm, "Inference attacks on location tracks," in *Proceedings of the International Conference on Pervasive Computing*. Springer, 2007, pp. 127–143.

[11] M. Shin, C. Cornelius, D. Peebles, A. Kapadia, D. Kotz, and N. Triandopoulos, "Anonysense: A system for anonymous opportunistic sensing," *Pervasive and Mobile Computing*, vol. 7, no. 1, pp. 16–30, February 2011.

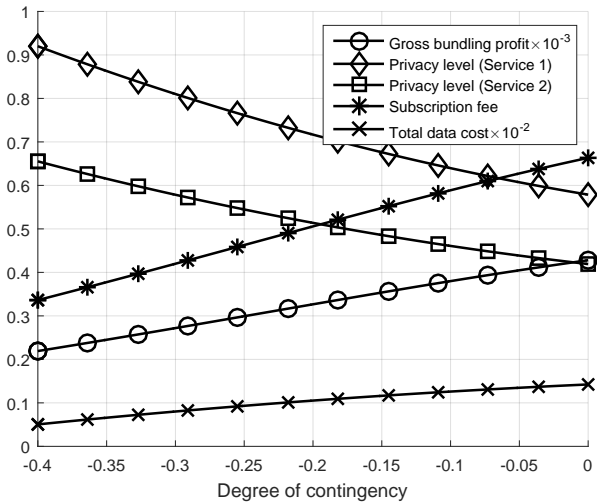[12] L. Pournajaf, L. Xiong, V. Sunderam, and S. Goryczka, "Spatial task

Fig. 18: Impacts of the degree of contingency γ on the gross profit $G_s^*(\cdot)$, privacy levels r_1^* and r_2^*, subscription fee p_b^*, and total data cost.
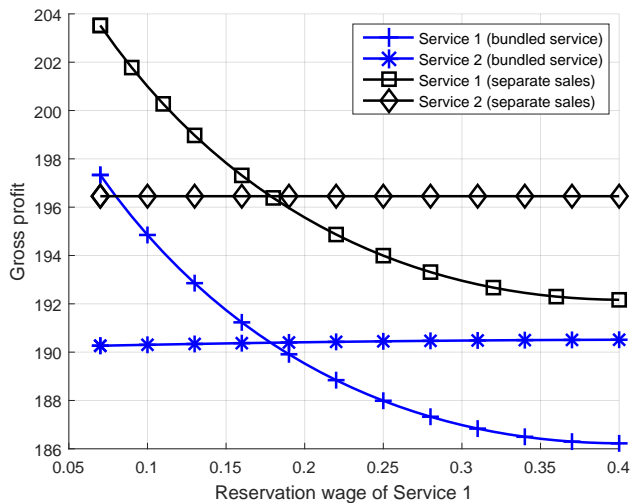

Fig. 19: Profits of S_1 and S_2 when sold separately or bundled as substitutes.

assignment for crowd sensing with cloaked locations," in *Proceedings of the 15th International Conference on Mobile Data Management*. IEEE, 2014, pp. 73–82.

[13] L. Sweeney, "k-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, October 2002.

[14] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 1, p. 3, March 2007.

[15] C. Dwork, "Differential privacy: A survey of results," in *Proceedings of the International Conference on Theory and Applications of Models of Computation*. Springer, 2008, pp. 1–19.

[16] N. Mohammed, R. Chen, B. Fung, and P. S. Yu, "Differentially private data release for data mining," in *Proceedings of the 17th SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2011, pp. 493–501.

[17] X. Xiao, G. Wang, and J. Gehrke, "Differential privacy via wavelet transforms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 8, pp. 1200–1214, August 2011.

[18] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the SIGSAC Conference on Computer and Communications Security*. ACM, 2016, pp. 308–318.

[19] L. Duan, T. Kubo, K. Sugiyama, J. Huang, T. Hasegawa, and J. Walrand, "Incentive mechanisms for smartphone collaboration in data acquisition and distributed computing," in *Proceedings of the The 31st International Conference on Computer Communications*. IEEE, 2012, pp. 1701–1709.

[20] Y. Wen, J. Shi, Q. Zhang, X. Tian, Z. Huang, H. Yu, Y. Cheng, and X. Shen, "Quality-driven auction-based incentive mechanism for mobile crowd sensing," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 9, pp. 4203–4214, September 2015.

[21] T. Luo, S. Kanhere, S. Das, and H. Tan, "Incentive mechanism design for heterogeneous crowdsourcing using all-pay contests," *IEEE Transactions on Mobile Computing*, vol. 15, no. 9, pp. 2234–2246, September 2016.

[22] D. Niyato, M. Abu Alsheikh, P. Wang, D. I. Kim, and Z. Han, "Market model and optimal pricing scheme of big data and Internet of things (IoT)," in *Proceedings of the 6th International Conference on Communications*, 2016, pp. 1–6.

[23] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," http://www.sentiment140.com, p. 12, 2009.

[24] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SigKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, December 2011.

[25] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[26] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, October 2001.

[27] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and Systems Magazine*, vol. 6, no. 3, pp. 21–45, March 2006.

[28] C. C. Aggarwal and S. Y. Philip, "A general survey of privacy-preserving data mining models and algorithms," in *Privacy-preserving data mining*. Springer, 2008, pp. 11–52.

[29] The Apache Foundation, "Apache Mahout: Scalable machine learning and data mining," http://mahout.apache.org, online; accessed January 2017.

[30] X. Meng, J. Bradley, B. Yavuz, E. Sparks, S. Venkataraman, D. Liu, J. Freeman, D. Tsai, M. Amde, S. Owen *et al.*, "MLlib: Machine learning in Apache Spark," *Journal of Machine Learning Research*, vol. 17, no. 34, pp. 1–7, April 2016.

[31] T. Strutz, *Data fitting and uncertainty: A practical introduction to weighted least squares and beyond*. Vieweg and Teubner, 2010.

[32] E. K. Chong and S. H. Zak, *An introduction to optimization*. John Wiley & Sons, 2013, vol. 76.

[33] R. Venkatesh and W. Kamakura, "Optimal bundling and pricing under a monopoly: Contrasting complements and substitutes from independently valued products," *The Journal of Business*, vol. 76, no. 2, pp. 211–231, April 2003.

[34] R. B. Myerson, *Game theory*. Harvard university press, 2013.

BIOGRAPHIES

Mohammad Abu Alsheikh [S'14] (mohammad027@e.ntu.edu.sg) received the B.Eng. degree in computer systems engineering from Birzeit University, Palestine, in 2011. Between 2010 and 2012, he was a Software Engineer working on developing robust web services, Ajax-based web components, and smartphone applications. He is currently a Ph.D. candidate in the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include machine learning in big data analytics, mobile sensing technologies, and sensor-based activity recognition.

Dusit Niyato [M'09–SM'15–F'17] (dniyato@ntu.edu.sg) is currently an Associate Professor in the School of Computer Science and Engineering, at Nanyang Technological University, Singapore. He received B.Eng. from King Mongkuts Institute of Technology Ladkrabang (KMITL), Thailand in 1999 and Ph.D. in Electrical and Computer Engineering from the University of Manitoba, Canada in 2008. His research interests are in the area of energy harvesting for wireless communication, Internet of Things (IoT) and sensor networks.

Derek Leong (dleong@i2r.a-star.edu.sg) received the B.S. degree in electrical and computer engineering from Carnegie Mellon University in 2005, and the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology in 2008 and 2013, respectively. He is a scientist with the Smart Energy and Environment cluster at the Institute for Infocomm Research (I^2R), A*STAR, Singapore. His research and development interests include distributed systems, sensor networks, smart cities, and the Internet of Things.

Ping Wang [M'08–SM'15] (wangping@ntu.edu.sg) received the Ph.D. degree in electrical engineering from University of Waterloo, Canada, in 2008. Currently she is an Associate Professor in the School of Computer Science and Engineering, Nanyang Technological University, Singapore. Her current research interests include resource allocation in multimedia wireless networks, cloud computing, and smart grid. She was a corecipient of the Best Paper Award from IEEE Wireless Communications and Networking Conference (WCNC) 2012 and IEEE International Conference on Communications (ICC) 2007.

Zhu Han [S'01–M'04–SM'09–F'14] (zhan2@uh.edu) received the B.S. degree in electronic engineering from Tsinghua University, in 1997, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 1999 and 2003, respectively. From 2000 to 2002, he was an R&D Engineer of JDSU, Germantown, Maryland. From 2003 to 2006, he was a Research Associate at the University of Maryland. From 2006 to 2008, he was an assistant professor in Boise State University, Idaho. Currently, he is a Professor in Electrical and Computer Engineering Department as well as Computer Science Department at the University of Houston, Texas. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, wireless multimedia, security, and smart grid communication. Dr. Han received an NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the EURASIP Best Paper Award for the Journal on Advances in Signal Processing in 2015, several best paper awards in IEEE conferences, and is currently an IEEE Communications Society Distinguished Lecturer.