IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Bi-Tier Differential Privacy for Precise Auction-based People-Centric IoT Service

Yuan Tian[1], Biao Song*[,2], Tinghuai Ma[2], Abdullah Al-Dhelaan[3], and Mohammed Al-Dhelaan[3]

[1] School of Computer Engineering, Nanjing Institute of Technology, Nanjing 210000, China
*[, 2] School of Computer Software, Nanjing University of Information Science and Technology, Nanjing 210044, China
[3] Dept. Computer Science, King Saud University, Riyadh, KSA

Corresponding author: Biao Song (e-mail: bsong@nuist.edu.cn).

This work was supported by Startup Foundation for Introducing Talent of Nanjing University of Information Science and Technology (Grant No.2019r030). The authors extend their appreciation to the Natural Science Foundation of the Jiangsu Higher Education Institute of China (20KJB413001), the Scientific Research Foundation of Nanjing Institute of Technology (YKJ201922), and the Deanship of Scientific Research at King Saud University for funding this work through research group no. RGP-264.

ABSTRACT With the fast proliferation of device sensing and computing, crowed sensing has become the building block of the Internet of things. Consequently, various data collection and incentive mechanisms are investigated for people-centric services. In this paper, we have investigated the problem of privacy-aware people-centric IoT service based on a tailored auction approach. We applied a bi-tier differential privacy methodology on the data collected from crowdsensing IoT devices. A corresponding pricing scheme is also proposed to ensure the property of incentive compatibility, precise service data, and anonymized query results. Comparing to traditional privacy-aware auction schemes which only focus on the cost, our corresponding precise privacy-aware auction scheme provides a tailored IoT service based on the customers' request. The proposed trial query technique is able to provide a precise assessment of service quality, thus improves the efficiency of the people-centric IoT service. The customer could enjoy the convenience of service evaluation before making a bid, while the actual service data is anonymized to guarantee the service providers' interests. We evaluate the proposed bi-tier differential privacy schema for auction-based service by conducting extensive simulations. The experimental results show that our proposed method yields higher data utility and accuracy for the IoT service customers with privacy concerns.

INDEX TERMS Data Protection, Internet of Things, Differential Privacy, Crowd Sensing IoT System

I. INTRODUCTION

Crowd sensing techniques emerge as a powerful solution in Internet of Things [1] The large amount of sensed data is aggregated and analyzed by the IoT service provider and people-centric services are provided for the customers to subscribe [2]. Such new paradigm, as the result, promotes extensive applications in the area of healthcare, anomaly detection and transportation [3-5, 37].

Despite the convenience of people-centric IoT service [30, 31,], the collection of sensing data suffers the dangerous from the privacy and security breaches [18]. Users have concerns when their data is stored, processed, and analyzed by the third party. Therefore, the technical challenges in the privacy-aware data collection need to be investigated.

Considerable state-of-the-art privacy protection research has been conducted for data collection [28, 29]. The most commonly used approach is the anonymous series models [6-9]. Samarati et al. [14, 15] introduced the concept of k-anonymity and numerous k-anonymity algorithms [16, 17] are then designed by generalization and suppression. l-diversity was proposed later [9] to overcome the limitation of k-anonymity by providing sufficient privacy protection against attribute disclosure. However, the above approach fails due to their high complexity, poor controllability, and dependence on the background knowledge of adversaries. Anonymization would not be enough if the attacker has auxiliary information from other data sources [13]. Differential privacy is then proposed to against adaptive attacks that use auxiliary information [14]. Differentially private algorithm incorporates random noise to make the data become noisy and imprecise.

Besides privacy-aware data collection, an incentive mechanism with service evaluation and the privacy protection of the service data are crucial to the quality of people-centric services. How to provide the customer a precise evaluation for the service he may subscript, while without compromising the service provider's real data is a challenging issue.

In this paper, we investigate the issue of privacy-aware people-centric IoT service subscription, and aim to provide a tailored auction mechanism for both customers and service providers with the properties of incentive compatibility, precise service data, and anonymized query results. We applied a bi-tier differential privacy methodology on the data collected from crowd sensing IoT devices. The IoT service providers evaluate the received anonymized data to build different types of human-centric

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3067138, IEEE Access

IoT services for the customers to subscribe. In order to help the customer get a more precise evaluation of quality of the data service, we allow the customer sends a trial query to see if the service data is meet his/her expectation. Since the exact query results contains valuable information and could not be released to the customers before they actually paid, a trial query result will be sent to the customer, which maintains the basic data pattern for service evaluation, and does not disclose the real data in the service. We apply the wavelet clustering method to preserve the service data privacy on the service provider's end, which is efficient in terms of time complexity and suitable for large special database.

Comparing to traditional privacy-aware auction schemes [32-35] which only focus on the cost, our corresponding precise privacy-aware auction scheme is proposed to provide customer's a tailored IoT service based on customer's request quality. The assessment improves the efficiency of the people-centric IoT service without unnecessary payment from the customer's side. Many people-centric IoT services with numeric data generated are in the scope of this work, such as location-based services, smart home, health care, etc.

The remainder of this paper is organized as follows. Section II discusses the related work. Section III presents the proposed system model and assumption scenario. Section IV and Section V describe the details of our incentive schemes. Section VI shows the experimental results and analyses. Section VII concludes the paper.

II. RELATED WORK

A. DIFFERENTIAL PRIVACY

The concept of privacy is defined as "the right to be left alone, free from intrusion or interruption" [38]. Privacy encompasses the categories of communication privacy, information privacy and physical privacy.

Differential privacy appears as a new notion of privacy with strong and mathematical in the field of statistical and machine learning analysis. It is a strictly provable and privacy controlled method and has become a de facto standard for a security-controlled privacy guarantee [39]. With the help of differential privacy, the modifications on a single database tuple would not affect the outcome of analysis.

The core idea of the differential privacy protection model can be classified into two categories: [40]. The first class is it ensures the operation of inserting or deleting a record in the input data set does not affect the output of any calculation (such as counting query). The second class is the model is not influenced by the attacker. With the background knowledge, even if the attacker has mastered sensitive information of all records except one record, the sensitive information of the record cannot be disclosed [41].

In local differential privacy method, the differentially private mechanism is applied to the real data on local device instead of sending them to the third-party aggregator. The aggregator receives the data that is already anonymized, and the aggregator will not have any access to real data [42, 43]. Local differential privacy method offers a great advantage that the third-party aggregator does not have to be fully trusted. The user's private date is safe even if the aggregator is malicious, which make the local differential privacy model well-suited to wider conditions where the third party is barely trusted.

The main difference between local differential privacy and proposed bi-tier differential privacy is the improved flexibility for the service provider and customer. Both customers and service providers are able to control the cost for privacy protection based on their own assessment of the value of data.

B. INCENTIVE MECHANISM FOR SERVICE AUCTION

Participatory sensing is a paradigm for data collection and collective knowledge information about a state or the condition of interest [45]. Auction-based incentive mechanisms [44-48] are widely studied accordingly for crowdsourcing and crowd sensing. The work in [45] addressed the issue of incentive mechanism in participatory sensing applications. The proposed approach is incentive for user participation and truthful cost declaration. In [19], the authors propose two privacy-aware incentive schemes to allow each mobile user earn credits by contributing data without leaking which data it has contributed, and ensure dishonest users cannot abuse the system to earn unlimited amount of credits. [20] designed an incentive scheme by performing data aggregation and perturbation with reliable workers. In [49], the authors designed a sealed bid second price auction to motivate user participation whereas the platform utility is neglected in the auction process. [50] presents a reverse auction based dynamic price incentive mechanism. The user could sell his/her sensed data with the claimed bid prices to the service provider. However, the authors did not consider the truthfulness when they design the mechanism. In [12], the author proposed a second pricing scheme to achieve incentive compatibility. The proposed VCG scheme, however, could not maximize the seller's income so the resulting revenue is away from the optimal solution.

III. THE PROPOSED SYSTEM MODEL

Fig. 1 demonstrates the proposed system model of people-centric service. Staring from the data source, IoT devices sense and generate data, perform light-weight local DP to eliminate the risk of privacy leakage, and then submit data to service providers. Some IoT devices are capable of running light-weight LDP such as smartphones. Other IoT gadgets send the privacy-sensitive data to local personal devices to perform LDP before delivering the data to service providers. Once the data have been purchased, the data owner will receive revenue from service providers. This mechanism can motivate more participants to join the people-centric service subscription.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3067138, IEEE Access

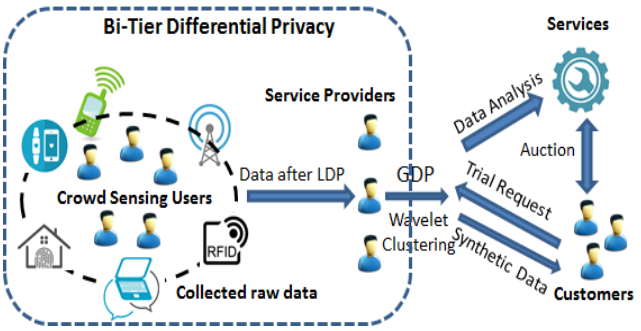

FIGURE 1. **The System Model of Bi-Tier Differential Privacy for Precise Auction-based People-Centric IoT Service.**

The service providers receive data from the data source, apply the GDP to further remove privacy risk. Meanwhile, the service providers may receive data queries from customers before they sign a contract. In order to provide precise data pattern information while keeping the real data secure, the service providers first adopt wavelet clustering to retrieve data pattern information, then generate synthetic data and send to customers.

Since the owners of IoT device adopt privacy protection technique, the data collected by service providers have different anonymization levels. According to [21], the quality of data and service is inversely proportional to the anonymization level. In the first stage of bi-tier DP, the higher the personal privacy level, the less true data the service providers can get from data owner. Similarly, the service provider can only provide lower quality of service with decreased data analytic quality if higher privacy level is adopted at the global stage.

For the global DP stage, the relation between privacy level and service quality is discussed in [21]. Let $u(\bullet)$ and r represent the service quality and the privacy level, respectively. The three empirical assumptions are: 1). $u(\bullet)$ is nonnegative; 2). $u(\bullet)$ is inversely proportional to $r \in [0,1]$; 3). $u(\bullet)$ is convex and decreases at an increasing rate over r. Then the relationship between the service quality and the privacy level is proposed in the following function:

$$u(r) = \alpha_1 - \alpha_2 \exp(\alpha_3 r) \tag{1}$$

where α_1, α_2 and α_3 are curve fitting parameters. These parameters can be obtained using data from real-world experiments by solving the following minimization problem [18]

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

$r^{(i)}$ and $\tau^{(i)}$ are the privacy level and the real-world service quality in the ith experiments, respectively, and a set of B experiments are performed to obtain these parameters.

Although the method in [21] can effectively predict the quality of data and service at the global stage, it cannot be applied to local stage due to two reasons: 1). local devices only have limited computational capacity which may not be capable of running a set of B experiments to obtain important parameters; 2). the data owners may not be willing to share the original data or the privacy protection level they adopted so that the service providers are not able to perform the prediction due to the lack of the privacy level r at local stage. As the final quality of data and service depends on the DP performed at both local and global stages, we argue that predicting the quality in our scenario is not feasible in practice. Thus, allowing the customers to use trail query and performing individual quality assessment have been recognized as a more realistic method which not only overcomes the aforementioned problem, but also provides better flexibility to the customers when evaluating the quality of data and service.

IV. BI-TIER DIFFERENTIAL PRIVACY

In this section, we first explain the definition of existing differential privacy technique, and then propose a bi-tier differential privacy solution.

A. DP PRELIMINARIES

The initial concept of DP was conceived to formalize the notion of privacy. The research work in [22] was practically achievable as well as theoretically verifiable. A comprehensive study of differential privacy is presented in [23] to provide an insight of DP from different research perspectives.

Consider sets D and D' that differ in maximum one data entry, let R be a randomized mechanism generating a query of a database with some probability. Let E be an event (a subset of query outcomes). We define the mechanism R is ε −differentially private (ε -DP) if for any E, D and D',

$$\Pr[R(D) \in E] \le e^{\varepsilon} \Pr[R(D') \in E] \tag{3}$$

In Eq. (1), the randomness of mechanism R leads to the probability. This inequality represents a worst-case privacy guarantee as this is a strong constraint that must hold for all E, and all random datasets D and D'.

Definition 1 (Differential privacy)

The notion of differential privacy (ε, δ) differential privacy is a more practical notion as a generalization of ε -DP. The explanation is that the mechanism R is (ε, δ) − differentially private (or (ε, δ) −DP in short) if for any E, D and D' (D and D' differs at most one entry),

$$\Pr[R(D) \in E] \le e^{\varepsilon} \Pr[R(D') \in E] + \delta \tag{4}$$

Given n is the input size, $\delta = \Omega(1/n)$ is recognized as poor privacy since private individual data can be fully

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3067138, IEEE Access

recovered [EYE-8]. A mechanism R can be (ε,δ) −DP for more than one pair of (ε,δ). Generally speaking, smaller (ε,δ) means better privacy.

B. LOCAL DIFFERENTIAL PRIVACY

In the proposed system, the service providers collect IoT data, reorganize and analyze the data, and publish their services. The privacy of users' data needs to be ensured during at two different stages. In the LDP model, we assume that the service provider has collected information about the identity of the user, but does not know the privacy information of each user. We set n as the total number of users, and a_i $(1 \le i \le n)$ is expressed as the i-th user. The privacy data of each user ai is represented by a tuple t_i, which contains c attributes A_1, $A_2 ... A_n$. These attributes can be numeric attributes or categorical attributes. In general, we assume that each numeric attribute has a field $[-1, 1]$, and each categorical attribute with a different k value has a discrete field$\{1, 2, \ldots, k\}$.

To protect privacy, each user first uses a random perturbation function f to perturb their tuples and then sends the perturbed data $f(t)$ instead of the original real data t to the aggregator or service provider. The perturbation function determines the trade-off between privacy security levels and data quality. If $f(t) = t$ is directly used, in this case, the user sends the real data directly to the service provider, and the service provider performs modeling calculation based on the real similar data, which will reach the extreme of the service quality, and at the same time This is also the fact that the aggregator completely receives the user's private data and will have no user privacy. If you let f output a tuple unrelated to t, the service provider will not perform computational modeling based on any valid user data, which also achieves extreme user privacy, but the quality of service will be reduced to zero.

Definition 2 (ε -local differential privacy)

The random function f satisfies $\varepsilon-$**local** differential privacy if and only if any two input tuples$t, t' \in Dom(f)$. At this point for any possible output$(t) = t^*$, we have:

$$\Pr[f(t) = t^*] \le e^{\epsilon} \times \Pr[f(t') = t^* \quad (5)$$

LDP is a special case of DP because random perturbation in LDP is performed by the user rather than by the aggregator, which means that the aggregator does not have the user's real private data. Due to the fact that local device may only have limited resources, the computational complexity and memory consumption of LDP need to be controlled strictly. We suggest that light weight perturbation methods should be used in this stage. Moreover, the size of data entry D need to be reduced, which can be achieved by restricting the scope of interest or sampling a subset of data from the original data entry. For example, if the data is related to location, the size of interested area can be restricted when forming the data entry.

According to the definition above, it can be ensured that the aggregator accepting the disturbed data tuple t^* cannot distinguish whether the true tuple is t or another tuple t' (controlled by the privacy parameterε). This also provides users with certain denial. At the same time, in the ε -local differential privacy, since the random disturbance is performed by the user himself, depending on the different privacy requirements of the users, different users further achieve the privacy protection by giving different values of the privacy parameter ε. The light weight implementation methods of LDP can be found in [42].

C. GLOBAL DIFFERENTIAL PRIVACY

The service providers apply global DP for two major reasons: 1). Protecting data owners' privacy; 2). Protecting service providers' valuable information. Since the service providers continuously collect large sets of IoT data, it is inevitable to see that some data owners do not protect their data properly. Meanwhile, the combination of data from different sources may result in new privacy issues. Thus, global DP will effectively protect data owners' privacy when such data is going to be delivered to customers as part of the service. On the other hand, certain valuable information may exist in the data set before or after the analysis enforced by service providers. To hide such information while remaining the service available, global DP technique is a promising solution due to the fact that the implementation of global DP is to add noise to the data set.

We consider Laplacian distribution and use it to achieve global DP. If Δf is the sensitivity of the function$f$, then we consider how the function may reveal the metric of the function, followed by adding Laplace noise to preserve $(\varepsilon, 0)$ differential privacy with a $\Delta f/\varepsilon$ scale.

Technically, Δf is I_1sensitivity. However, the results of Gaussian noise involve I_2 sensitivity. The main reason for this relies on the norm we use to measure the sensitivity (we use the sum of absolute values rather than the root mean square). The Laplace mechanism implements differential privacy by adding random noise that conforms to the Laplacian random distribution to the determined query results. The probability density function is:

$$f(x|\lambda) = (1/2\lambda)\, exp(-|x|/\lambda) \quad (6)$$

Where $\lambda > 0$ is the scale parameter of the variable x, which is determined by the sensitivity Δf of the function f and the privacy parameterε, $\lambda = \Delta f/\varepsilon$.

When we want to issue a function f: $\mathrm{D} \to \mathbb{R}^{\mathrm{d}}$, for the statistical data set, the closer the published data set and f(D), the better the privacy protection effect, and the privacy protection algorithm can be designed by the Laplace mechanism.

V. AUCTION-BASED SERVICE SUBSCRIPTION

In this section, we first introduce the wavelet clustering and synthetic data generation techniques that allow the service providers to answer the trail queries with unreal but useful

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3067138, IEEE Access

contents. Then the auction model for service subscription is introduced.

A. ANSWERING THE TRAIL QUERIES

When the original reply for a trail query is ready, the service provider will treat the multidimensional data as a multidimensional signal. It first divides the data space into a grid structure, and then transforms the data space into a frequency domain space by wavelet transform. After convolution with a kernel function in the frequency domain space, the natural clustering property [36] of the data appears. The wavelet clustering method is a multi-resolution algorithm with high resolution for detailed information and low resolution for contour information. We must mention that service providers will not be rewarded immediately when answering a trail query. Consequently, we choose wavelet clustering for the following reasons: 1) the time complexity of wavelet clustering is $O(n)$; 2) the wavelet clustering performs well on big multi-dimensional data. Without much overhead, the service providers can easily handle the trail queries.

Let n be the common length of the p series individually denoted by$X(i)$, which are the original data. At the beginning, given an orthogonal wavelet ψ, to decompose the aforementioned series at level J :

$$X^{(i)} = A_J^{(i)} + \sum_{j=1}^{J} D_j^{(i)} \tag{7}$$

where $A_J^{(i)}$ and $D_j^{(i)}$ are respectively the approximation and detail at level j of the signal $X^{(i)}$. Thus, we can say that$A_J^{(i)}$ and $D_j^{(i)}$ is in the spaces $V_j$and W_j respectively. The space W_j is the orthogonal complement of V_j intoV_{j-1}, $\left(V_{j-1} = V_j \oplus W_j\right)$. Consequently, the signal in V_j can be treated as approximation signals like A_j or equivalently denoted by cA_j in level-dependent bases.

The main idea of wavelet clustering is to first quantize the original data set into the feature space, perform wavelet transform on the feature space, and find the connected parts in the space after wavelet transform, which is clustering, not every cluster. The tag is added, and then the mapping representation provided by the algorithm determines the cluster to which each data point in the original data set belongs, so that the service providers can take appropriate data points from each cluster to form a model which can be used to generate synthetic data to answer the query.

We use the definition found in [24] to denote cluster(X) as the Ward hierarchical clustering algorithm applied to the matrix of Euclidean distances between objects described by X. The procedure of wavelet clustering is described as follows:

1.Decompose the signal with a signal waveletA_j , generate$\left(cA_J, cD_J, \ldots, cD_1\right)$.

2.Use various wavelet bases and calculate a set of partitions denoted as$clusters\left(cA_J, cD_J, \ldots, cD_1\right)$.

3.Choose the optimal partition by selecting a level of decomposition and a number of clusters C.

After finding the cluster information, the service provider is now able to generate synthetic data accordingly. In [25], we found the method to generate random data using wavelets. Given a sample of size $M = 2^K$ where K is a positive integer, taken from a stochastic process $f(t)$ with zero mean: $f_1, f_2, \ldots, f_M$ where $f(t) = \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} d(k,l)\, 2^{-k/2}\psi(2^{-k}t - l)$ and $d(k,l)$ is the discrete wavelet transform $d(k,l) = \int_{-\infty}^{\infty} f(t)2^{-k/2}\psi(2^{-k}t - l)dt$. Define the sample $f_{ki}(k = 0,1,\ldots,K; i = 1,\ldots,M)$ consisting of averages of 2^k successive elements of the sample. f_{0i} is the original sample and f_{Ki} is a sample of all zeros, since the average of M elements is zero. The detail function $g_k(t)$ has a sample made up of M elements $g_k(t) = f_{k-1}(t) - f_k(t)$ for $k = 1,2,\ldots,K$. Thus we have, for each element f_i of the original sample, K detail function values g_{ki} corresponding to di erent resolutions. Choosing randomly from among M elements for each $g_k(t)$, and then summing them up by $f(t) = \sum_{-\infty}^{\infty} g_k(t)$ we get a simulated value for $f(t)$ as $f_j = \sum_{k=1}^{K} g_{kt}$ where j is the index for generated elements.

B. AUCTION MODEL

For the simplicity of description, we assume one service provider providing S types of privacy-ware services to N customers. The services are distinguished by the functions, and each service owns P types of QoS (quality of service) levels corresponding to P types of global privacy protection level. Based on the type and level of services, we propose an auction model for the customers to submit bid for combinatorial services. It is worth to mention that the users are able to assess the quality of any service with trial query prior before submitting bid.

In the auction model, we consider the service provider as a single seller and the customers as multiple buyers. Regarding the auctioneer, the service provider can also run the tasks of starting auctions, collecting bid submissions, determining winners and final prices. For example, the service provider announces a group of privacy-aware services to customers. Let $SV_1, SV_2, \ldots, SV_S$ be S different services provided by the service provider. In order to manage the computation and memory resources, we assume that each service SV_i is associated with computational overhead C_i , network bandwidth B_i and memory overheadM_i. To be more specific, the P levels of services $SV_{i1}, SV_{i2}, \ldots, SV_{iP}$ are provided corresponding to varying privacy levels of people-centric data.

After knowing the announcement of privacy-aware services, customers will first submit trail queries to evaluate the quality of service for those they are interested in. Based on the evaluation results, they can estimate price and upload bids on their desired services to the service provider. We assume that that all customers are single-minded, each of them only creates a bid for one service bundle in each round. This assumption is rational and widely mentioned in auction models.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3067138, IEEE Access

Given a customerCR_i, we define the bid of CR_i by the following

$$B_i = \{d_0^i, d_1^i, \ldots, d_S^i, e_i\} \quad (8)$$

where d_0^i and r_i are a vector expressing buyer i's demand on service SV_i and the buyer i's valuation of requested service bundle, respectively. The demand d_0^i can be further expressed as

$$d_j^i = \{a_{j1}^i, a_{j2}^i, \ldots, a_{jP}^i\}, \forall j \in [0, S] \quad (9)$$

where $a_{jk}^i \in \{0,1\}$ represents whether the kth level of service of SV_j is required by customers i.

After bid collection from all the buyers, the auctioneer determines the winners and final prices of services to charge winners, based on a designed auction mechanism. These winners will make payment and receive services in their bid service bundle from the service provider.

VI. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we provide several simulations to evaluate our proposed methods on the empirical smart home dataset. For the bi-tier differential privacy solution, we set both informed and blind attack for testing, and then examine the service provider's ability and the external attacker's ability to classify the information. For the service subscription solution, we let service provider offer data filtering service, and assume customers use the data for further information retrieval. The effectiveness of trial query solution is examined followed by the evaluation of auction scheme in terms of average revenue and customers' satisfaction.

A. BI-TIER DP

In this part, we choose the dataset found in [26] for smart home traffic gathered from a real-world deployment of IoT devices. 30 IoT devices have been observed over a period of two weeks to form the dataset. To simulate the attacker's activity, we study two models.

In informed attack, we assume that the attacker know the exact number and type of devices used in a smart home, and perform the mapping between known device list to data source in order to infer the user's activities. This attack is more likely to happen when service provider is not trustful or not secured (only LDP has been applied). In blind attack, we believe that the attacker does not have the number and type of devices installed in the smart home, but still want to classify the data to identify the data source. We believe the external attackers can easily perform this task (bi-tier DP has been applied).

Another parameter we must mention is the ε value for ε -differential privacy. Let $S_{l_1}(Q)$ of a query set Q be l_1 sensitivity, we need to add Laplace noise with mean 0 and magnitude λ to the query results such that $S_{l_1}(Q) \leq \varepsilon\lambda$. With fixed sensitivity, larger values of ε denotes adding less noise and consequently lower levels of privacy protection. The value of ε for LDP in our experiments is in the range of [0.2, 1]. With lower computational capability, each data owner is given a randomly selected dataset of 10% size and randomly chooses a ε number from this range to perform LDP. For the GDP, the default value of ε is 0.3 with the entire dataset as the base of DP.

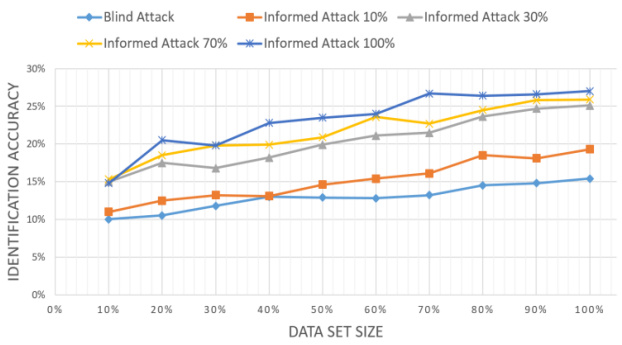

FIGURE 2. **Data set size vs. identification accuracy.**

Figure 2 shows the results for privacy protection against both informed and blind attacks on the data sets with different size. With the increasing percentage of information obtained by the attacker, the chance of identify data owner increases by more than 10% for the informed attacker model as compare to the blind attacker. When the attacker gets more than 70% of information, the privacy risk raises up to 25% or above under current settings. However, the blind attacker can only achieve 15% of accuracy when bi-tier DP is fully implemented. These results prove the effectiveness of GDP.

Figure 3 reports the identification accuracy as a function of increasing number of attributes. In the experiments, an attribute set of size within [1, 10] consists of the first n elements of the data set we used. Our results show that the privacy risk does not vary much with the number of attributes as long as the number of attribute is greater than 2. The reason behind this is that 3 or more attributes are enough for the attacker to perform identification task. For 8, 9 and 10 attributes, the informed attacks with more than 70% of information perform similarly. As expected, the blind attack shows less risk than the informed attack.

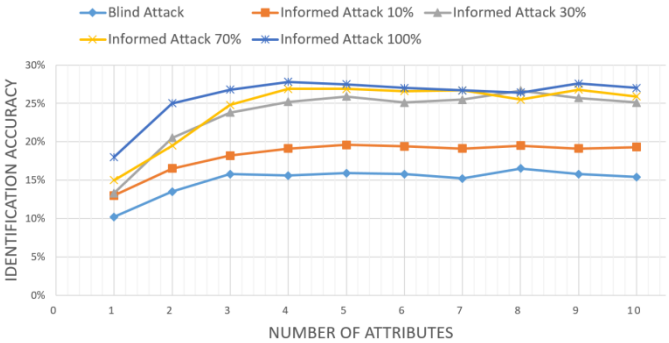

FIGURE 3. **Number of attribute vs. identification accuracy.**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3067138, IEEE Access

The parameter ε is inversely related to the privacy guarantees of differential privacy. In other words, at lower values ofε, Laplace noise of larger magnitude is added to query responses. Consequently one would expect less efficient blocking as ε declines. This finding is also supported by Figure 4. Obviously, the blind attack is more sensitive to the change of ε.

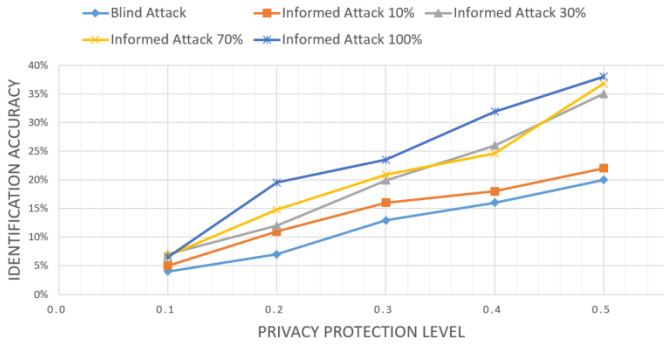

FIGURE 4. Privacy protection level vs. identification accuracy.

Figure 5 shows the comparison results of computational overhead at different stages. For this purpose, we use default settings and run both LDP and GDP on the same computer. It can be seen that the overhead of LDP is significantly lower than the overhead of GDP. Considering the lower computational capacity of local device, we can see the benefit of LDP. Besides, the overhead of GDP decreases when we relax the privacy protection requirement. This happens because adding noise in GDP is a heavy task, which only causes trivial overhead in LDP.

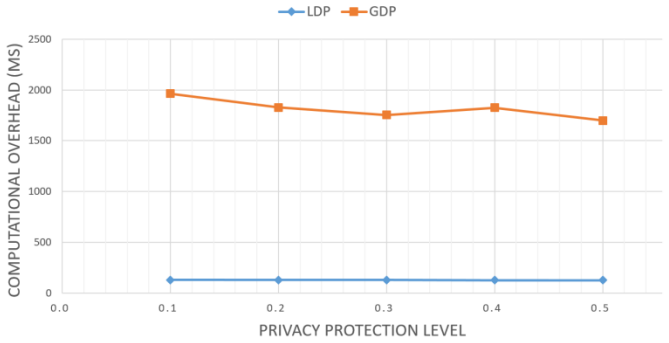

FIGURE 5. Privacy Protection level vs. computational overhead.

B. SERVICE SUBSCRIPTION

In the following simulations, we first examine the impact of bi-tier DP on the quality of data for both trial query and actual query. Assume that each customer is interested in several value ranges. If a data point falls into the same range before and after applying DP, we say the data point is valid. Otherwise, the data point is considered as invalid one. The quality of data is measure by the number of valid data points divided by the number of invalid data points. The rest of experiment settings remain default.

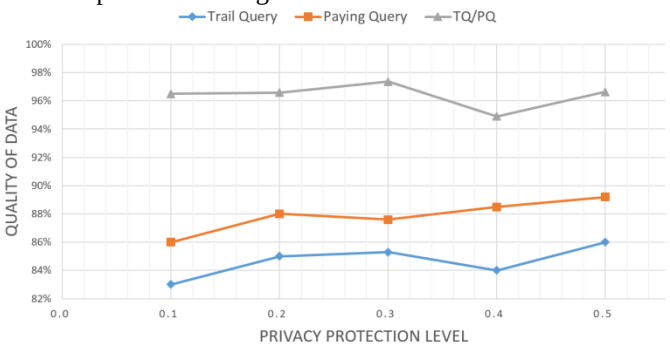

FIGURE 6. Privacy Protection level vs. quality of data.

Figure 6 shows the results of data quality affected by DP protection for both trail query and paying query. Remind that paying query deliveries data set after bi-tier DP, and trail query generate synthetic data set using wavelet clustering technique. With the decreasing privacy protection level, the quality of data slightly increases by less than 5%. Note that these results are conducted under current service type and system settings. The privacy protection level may have larger impact on the quality of data in a different scenario. The effectiveness of proposed trail query method can be seen from the same results. Although it is synthetic data, thanks to the wavelet clustering and random number generator, the quality of trail query data can keep up to 97% of data quality as compare to the output of paying query. But the exact data points are generated randomly.

Figure 7 reports the computational overhead as a function of increasing data set size. In the experiments, the size of data set is increased from 20% to 100%. Our results show that the computational overhead for handling both type of query increases linearly. For the paying query, we get this result since the service provider is providing filtering service with computational complexity *O(n)*. For the trail query, the additional overhead is from the wavelet transform and the random number generation, where both tasks have the same computational complexity *O(n)*. Note that the frequency of answering trail query is much lower than the frequency of answering paying query. Thus, the additional overhead for handling trial query is acceptable.

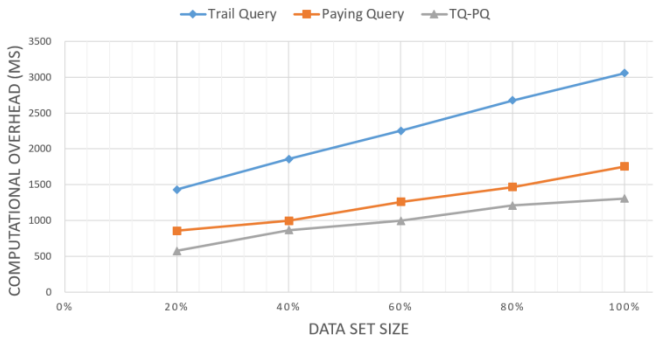

FIGURE 7. Data set size vs. computational overhead.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3067138, IEEE Access

In last simulation, we limit the capacity of service provider and examine the customer satisfaction rate. Specifically, if a customer's requested service bundle is allocated, the customer is satisfied. The combinatorial auction bid may not be fulfilled by the service providers due to the lack of resources, thus making the customer unsatisfied.

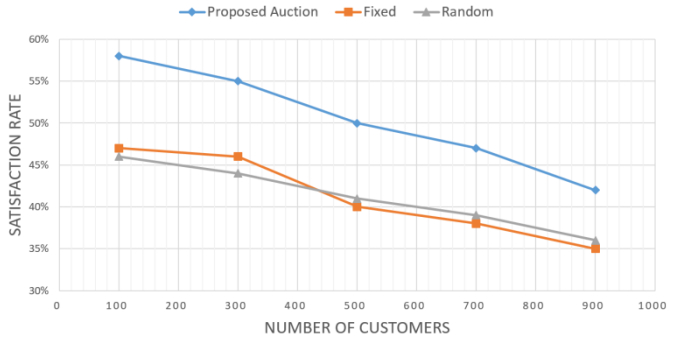

FIGURE 8. Number of customers vs. Satisfaction rate.

We compare the results with a fixed allocation scheme [27] and a random allocation scheme [27] in Figure 8. As we can see that for these mechanisms, user satisfaction rate decreases as the number of users increases. In addition, we observe that the proposed solution provides better satisfaction than the other methods. Our solution outperforms the fixed and random schemes because the service provider in those schemas offers limited resource combinations which often fail to meet the varying requirements of customers.

VII. CONCLUTIONS

In this paper, we address the major technical challenges privacy-aware service subscription in people-centric IoT sensing, and validate its operation and effectiveness with the prototype system. We applied a bi-tier differential privacy methodology on the data collected from crowd sensing IoT devices. A corresponding pricing scheme is also proposed to ensure the property of incentive compatibility, precise service data, and anonymized query results. Comparing to traditional privacy-aware auction schemes which only focus on the cost, our corresponding precise privacy-aware auction scheme is proposed to provide customer's a tailored IoT service based on customer's request quality. The assessment improves the efficiency of the people-centric IoT service without unnecessary payment from the customer's side. Numerical analyses show the effectiveness of our proposed methodology.

ACKNOWLEDGMENT

This work was supported by Startup Foundation for Introducing Talent of Nanjing University of Information Science and Technology (Grant No.2019r030). The authors extend their appreciation to the Natural Science Foundation of the Jiangsu Higher Education Institute of China (20KJB413001), the Scientific Research Foundation of Nanjing Institute of Technology (YKJ201922), and the Deanship of Scientific Research at King Saud University for funding this work through research group no. RGP-264.

REFERENCES

[1] S. Song, S. Shin, Y. Jang, S. Lee and B. Choi, "Effective Opportunistic Crowd Sensing IoT System for Restoring Missing Objects," 2015 IEEE International Conference on Services Computing, New York, NY, 2015, pp. 293-300.
[2] B. Guo, Z. Wang, Z. Yu, Y. Wang, N. Y. Yen, R. Huang, and X. Zhou, "Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm," ACM Comput. Surv., vol. 48, no. 1, pp. 1–31, Aug. 2015.
[3] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K. Kwak, "The Internet of Things for health care: A comprehensive survey," IEEE Access, vol. 3, pp. 678–708, 2015.
[4] M. Antonini, M. Vecchio, F. Antonelli, P. Ducange, and C. Perera, "Smart audio sensors in the Internet of Things edge for anomaly detection," IEEE Access, vol. 6, pp. 67 594–67 610, 2018.
[5] T. N. Pham, M.-F. Tsai, D. B. Nguyen, C.-R. Dow, and D.-J. Deng, "A cloud-based smart-parking system based on Internet-of-Things technologies," IEEE Access, vol. 3, pp. 1581–1591, 2015.
[6] T. Ma, Y. Zhang, J. Cao, J. Shen, M. Tang, Y. Tian, KDVEM: A k-degree anonymity with Vertex and Edge Modification algorithm, Computing, 97 (2015): 1165–1184.
[7] Y. Tian, M. M. Kaleemullah, M. A. Rodhaan, B. Song, A. Al-Dhelaan, T. M, A privacy preserving location service for cloud-of-things system, J. Parallel Distrib. Comput., 123 (2019), 215–222.
[8] 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, 2002.
[9] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," in Proceedinngs of 22nd IEEE International Conference on Data Engineering (ICDE), 2006, pp. 24–24.
[10] M. Du, K. Wang, X. Liu, S. Guo and Y. Zhang, "A Differential Privacy-Based Query Model for Sustainable Fog Data Centers," in IEEE Transactions on Sustainable Computing, vol. 4, no. 2, pp. 145-155, 1 April-June 2019.
[11] Cynthia Dwork. Differential privacy: A survey of results. International Conference on Theory and Applications of Models of Computation, 2008.
[12] L. Gao, P. Li, Z. Pan, N. Liu, and X. You, "Virtualization framework and VCG based resource block allocation scheme for LTE virtualization," in Proceedings of 2016 IEEE 83rd Vehicular Technology Conference (VTC), 2016, pp. 1–6.
[13] J. Ni, K. Zhang, X. Lin, and X. S. Shen, "Securing Fog Computing for Internet of Things Applications: Challenges and Solutions," IEEE Communications Surveys & Tutorials, vol. 20, no. 1, pp. 601–628, 2018.
[14] P. Samarati. Protecting respondent's privacy in microdata release. IEEE T. Knowl. Data En., 13(6):1010–1027, 2001. [16]
[15] P. Samarati and L. Sweeney. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical Report SRI-CSL-98-04, SRI Computer Science Laboratory, 1998
[16] R. J. Bayardo and R. Agrawal. Data privacy through optimal k-anonymization. In Proc. 21st Intnl. Conf. Data Engg. (ICDE), pages 217–228, Washington, DC, USA, 2005. IEEE Computer Society
[17] L. Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. Int. J. Uncertain. Fuzz., 10(6):571–588, 2002
[18] G. T. Duncan, S. E. Fienberg, R. Krishnan, R. Padman, and S. F. Roehrig. Disclosure limitation methods and information loss for tabular data. In Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies, pages 135–166. Elsevier, 2001.
[19] Q. Li and G. Cao, "Providing privacy-aware incentives in mobile sensing systems," IEEE Transactions on Mobile Computing, vol. 15, no. 6, pp. 1485–1498, 2015.

[20] Computer Networks, vol. 102, pp. 157–171, 2016. [14] H. Jin, L. Su, H. Xiao, and K. Nahrstedt, "Incentive mechanism for privacy-aware data aggregation in mobile crowd sensing systems," IEEE/ACM Transactions on Networking (TON), vol. 26, no. 5, pp. 2019–2032, 2018.
[21] M. A. Alsheikh, D. Niyato, D. Leong, P. Wang, and Z. Han, "Privacy management and optimal pricing in people-centric sensing," IEEE Journal on Selected Areas in Communications, vol. 35, no. 4, pp. 906–920, 2018.
[22] Cynthia Dwork. A firm foundation for private data analysis. Communications of ACM, 54(1):86 – 95, 2011.
[23] Cynthia Dwork. Differential privacy: A survey of results. In International Conference on Theory and Applications of Models of Computation, pages 1 – 19. Springer, 2008.
[24] Kaufman, Leonard, and Peter J. Rousseeuw. Finding groups in data: an introduction to cluster analysis. Vol. 344. John Wiley & Sons, 2009.
[25] Bayazit, Mehmetcik, and Hafzullah Aksoy. "Using wavelets for data generation." Journal of Applied Statistics 28, no. 2 (2001): 157-166.
[26] "Unsw iot smart home dataset." [Online]. Available: https://iotanalytics.unsw.edu.au/
[27] X. Zhang, H. Qian, K. Zhu, R. Wang, and Y. Zhang, "Virtualization of 5g cellular networks: A combinatorial double auction approach," in Proceedings of IEEE Global Communications Conference (GLOBECOM), 2017, pp. 1–6.
[28] Yang Y, Zheng X, Guo W, et al. Privacy-preserving fusion of IoT and big data for e-health[J]. Future Generation Computer Systems, 2018, 86:1437-1455.
[29] Yang Y, Zheng X, Guo W, et al., Privacy-preserving Smart IoT-based Healthcare Big Data Storage and Self-adaptive Access Control System[J]. Information Sciences, 2018, 479: 567-592.
[30] Chen X, Li A, Zeng X, et al. Runtime model based approach to IoT application development[J]. Frontiers of Computer Science, 2015, 9(4): 540-553.
[31] Yang Y, Zheng X, Tang C. Lightweight distributed secure data management system for health internet of things[J]. Journal of Network and Computer Applications, 2017, 89: 26-37.
[32] Guo W, Chen J, Chen G, et al. Trust dynamic task allocation algorithm with Nash equilibrium for heterogeneous wireless sensor network[J]. Security and Communication Networks, 2015, 8(10): 1865-1877.
[33] Chen X, Lin J, Ma Y, et al. Self-adaptive resource allocation for cloud-based software services based on progressive QoS prediction model[J]. Science China(Information Sciences), 2019, 62(11): 219101.
[34] Chen X, Wang H, Ma Y, Zheng X, Guo L. Self-adaptive resource allocation for cloud-based software services based on iterative QoS prediction model[J]. Future Generation Computer Systems, 2020, 105: 287-296.
[35] Chen X, Lin J, Lin B, et al. Self-learning and self-adaptive resource allocation for cloud-based software services[J]. Concurrency and Computation: Practice and Experience[J]. 2019, 31(23): e4463.
[36] He Z, Yu C. Clustering stability-based Evolutionary K-Means[J]. Soft Computing, 2019, 23(1): 305-321.
[37] Espada, Jordán Pascual, Yager R, Yu Z . Communications, collaborations and services in networks of embedded devices[J]. Future Generation Computer Systems, 2019, 92:560-563.
[38] International Organization for Standardization Information Technology—Business Operational View—Part 1: Operational Aspects of Open-Edi for Implementation20112ndLondon, UKInternational Organization for StandardizationISO/IEC 15944-1.
[39] Yuan, G., Zhang, Z., Winslett, M., Xiao, X., Yang, Y., Hao, Z.Low-rank mechanism: optimizing batch queries under differential privacy, Proceedings of the VLDB
[40] C.Dwork, "Differential Privacy," in Proc. of 33rd International Colloquium on Automata, Languages and Programming-ICALP 2006, Italy, 2006.
[41] R. Bassily and A. Smith, "Local, Private, Efficient Protocols for Succinct Histograms," in STOC, 2015.
[42] Wang, T., Zhang, X., Feng, J. and Yang, X., 2020. A Comprehensive Survey on Local Differential Privacy toward Data Statistics and Analysis. Sensors, 20(24), p.7030.
[43] H. Zhang, B. Liu, H. Susanto, G. Xue and T. Sun, "Incentive mechanism for proximity-based Mobile Crowd Service systems," IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications, San Francisco, CA, 2016, pp. 1-9.
[44] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing," in ACM MobiCom, 2012.
[45] I. Koutsopoulos, "Optimal incentive-driven design of participatory sensing systems," in INFOCOM 2013.
[46] D. Peng, F. Wu, and G. Chen, "Pay as How Well You Do: A Quality Based Incentive Mechanism for Crowdsensing," in ACM MobiHoc 2015.
[47] Y. Wei, Y. Zhu, H. Zhu, Q. Zhang, and G. Xue, "Truthful Online Double Auctions for Dynamic Mobile Crowdsourcing," in IEEE Infocom 2015.
[48] D. Zhao and X.-Y. Li and H. Ma, "How to crowdsource tasks truthfully without sacrificing utility: Online incentive mechanisms with budget constraint," in IEEE Infocom 2014.
[49] G. Danezis, S. Lewis, and R. Anderson. How much is location privacy worth? In Proceedings of WEIS, 2005.
[50] J. Lee and B. Hoh. Sell your experiences: A market mechanism based incentive for participatory sensing. In Proceedings of IEEE PerCom, pages 60–68, 2010.

BIOGRAPHY

Yuan Tian has received her master and Ph.D degree from KyungHee University in South Korea. After that, she joined King Saud University as an assistant professor in the College of Computer and Information Sciences. She is now an associate professor in Nanjing Institute of Technology. She is currently on the editorial boards of several journals, and has been the workshop/session chairs, organization and program committee for several reputable international conferences. She has participated more than ten national and industrial projects in Korea and Saudi Arabia such as National IT industry and National Research Foundation. Besides, she also works as PI and Co-PI in several projects including National Plan for Science, Technology and Innovation (NPSTI). Her main research interest is information protection and privacy preservation in the area of IoT, Location-based Service(LBS), Social Networks, Cloud Computing and Healthcare domain.

.

Biao Song received his Ph.D. degree in Computer Engineering from Kyung Hee University, South Korea in 2012. He has worked in King Saud University, Kingdom of Saudi Arabia as Assistant Professor since 2013, and now he is working in Nanjing University of Information Science and Technology in China. His current research interests are security , Cloud computing, remote display technologies and dynamic VM resource allocation.

Tinghuai Ma received the bachelor's and master's degrees from the Huazhong University of Science and Technology, China, in 1997 and 2000, respectively, and the Ph.D. degree from the Chinese Academy of Science, in 2003. He was a Post-Doctoral Associate with AJOU University, in 2004. From 2007 to 2008, he visited Chinese Meteorology Administration. In 2009, he was a Visiting Professor with the Ubiquitous Computing Laboratory, Kyung Hee University. He is currently a Professor in Computer Sciences with the Nanjing University of Information Science and Technology, China. He has authored over 100 journal/conference papers. His research interests are data mining, cloud computing, ubiquitous computing, and privacy preserving, and so on.

Abdullah Al-Dhelaan, has received BS in Statistics (Hon) from King Saud University, on 1982, and the MS and Ph.D. in Computer Science from Oregon State University on 1986

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3067138, IEEE Access

and 1989 respectively. He is currently the Vice Dean for Academic Affairs, Deanship of Graduate Studies and a Professor of Computer Science, King Saud University, Riyadh, Saudi Arabia. He has guest edited several special issues for the Telecommunication Journal (Springer), and the International Journal for Computers and their applications (ISCA). Moreover, he is currently on the editorial boards of several journals and the organizing committees for several reputable international conferences. His current research interest includes: Mobile Ad Hoc Networks, Sensor Networks, Cognitive Networks, Network Security, Image Processing, and High Performance Computing.

Mohammed Al-Dhelaan received his M.S. degree (2009) and his Ph.D. degree (2015) in Computer Science from The George Washington University. He, then, moved on to become an Assistant Professor in the Computer Science Department at King Saud University. His research interest includes Natural Language Processing and Data Mining. Specifically, he is focused on graph-based ranking, extracting keyphrases, statistical topic models, and text summarization. He has supervised M.S. students and taught several courses in both undergraduate and graduate levels. Also, he has been participating as a PC member in many conferences and have reviewed several research articles in the past years."