IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 26, 2017, accepted August 29, 2017, date of publication September 4, 2017, date of current version September 27, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2748956

PCP: A Privacy-Preserving Content-Based Publish–Subscribe Scheme With Differential Privacy in Fog Computing

QIXU WANG[1], DAJIANG CHEN[1,2], (Member, IEEE), NING ZHANG[3], (Member, IEEE), ZHE DING[1], AND ZHIGUANG QIN[1], (Member, IEEE)

[1]School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[2]Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[3]Department of Computing Science, Texas A&M University-Corpus Christi, Corpus Christi, TX 78412, USA

Corresponding author: Zhiguang Qin (zgqin@uestc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61133016, Grant 61472064, and Grant 61502085, in part by the National High-Tech Research and Development Program of China (863 Program) under Grant 2015AA016007, in part by the Projects of International Cooperation and Exchanges NSFC under Grant 61520106007, and in part by the China Post-Doctoral Science Foundation funded project under Grant 2015M570775.

ABSTRACT Fog computing dramatically extends the cloud computing to the edge of the network and admirably solves the problem that the brokers (in publish–subscribe system) generally lack of computing capacity and energy power. However, brokers may be disguised, hacked, sniffed, and corrupted. The traditional security technology cannot protect the system privacy when facing a possible collusion attack. In this paper, we propose a privacy-preserving content-based publish/subscribe scheme with differential privacy in fog computing context, named PCP, where the fog nodes act as the brokers. Specifically, PCP firstly utilizes the U-Apriori algorithm to mine the top-K frequent itemsets (i.e., the attributes) from uncertain data sets, then applies the exponential and Laplace mechanism to ensure the differential privacy, and the broker uses the mined top-K itemsets to match appropriate publisher and subscriber finally. Security analysis shows that the PCP can guarantee differential privacy in theory. To evaluate the performance of PCP, we carry out experiments with real-world scenario data sets. The experimental results show that PCP efficiently achieves the tradeoff between the system cost and the privacy demand.

INDEX TERMS Publish–subscribe, differential privacy, fog computing, privacy-preserving, uncertain datasets.

I. INTRODUCTION

The publish–subscribe (PS) system has been widely applied in many modern large-scale mission critical systems (LSMCS), such as manufacturing production and highway traffic monitoring [1]. As an important part of the PS system, the broker can be used to decouple the users interaction and provide asynchronous communications in LSMCS [1]. The broker communicates with different entities (e.g., publishers and subscribers), matches the suitable user requirement, and transmits users' data [2]. However, with the proliferation of mobile services and applications, brokers are required to equip with more computing capacity and energy power. The fog computing can dramatically address this issue well by extending the cloud computing to the edge of the network [3]. Fog computing immensely extends the terminal devices capacity and feasibility. Nevertheless, it also incurs much higher risks. For instance, there may exist an unethical broker. What is worse, the legitimate broker may face the challenges (such as hacking, sniffing and corrupting) from the potential malicious adversaries [4] and leak the privacy information of users. These security challenges make the broker become the vulnerability in fog-based PS system. How to protect the privacy of users becomes increasingly important.

The straightforward method to resolving this privacy problem is employing the cryptography mechanism [5]. The most common cryptography method is encrypting both publishers and subscribers datasets before sending these data to the helpers [4], [5]. It can protect the confidentiality and privacy of system users by directly using encryption algorithms. However, these traditional security solutions cannot prevent

the collusion attack. A malicious user (publisher or subscriber) who is supposed to keep the secret (the encryption key or the content of data) of other users would deliberately leak the secret to the hostile brokers. The malicious brokers and users collude with each other and share secret, e.g., 1) the malicious user provides other users' sensitive data to the collusive broker in support of analyzing these data, and 2) the malicious broker provides other users' data to its colluders so that the colluders could pretend as the most suitable candidate to other users. These security risks and vulnerabilities obstruct the wide deployment of fog-based PS system.

Differential privacy technology has the great potential to ensure the data privacy by preventing adversaries from analyzing data [6], [7]. It adds the artificial noise to the data before output it so that the adversary cannot figure out the actual data through the statistical analysis [7], [8]. This feature can exactly prevent the collusion attack well in the fog-based PS system. However, the brokers also need to match the same or the most similar interest between the publisher and subscriber in PS system. How to balance the privacy of users and the necessary matching information of personality, and protect the PS system from the collusion attack simultaneously, become an urgent and necessary issue.

In this paper, we propose a privacy-preserving content-based PS scheme with differential privacy in fog computing context (PCP). The PCP can ensure the privacy of users, the functions of PS system, and resist the collusion attacks. Specifically, the process of this scheme can be mainly divided into three phases. Firstly, the notification events are generated for all users, the publishers and subscribers, by leveraging the U-Apriori algorithm to mine the top-K frequent attributes (i.e., itemsets) from uncertain datasets, and applying the exponential mechanism to ensure the differential privacy in the mining step. Secondly, the Laplace mechanism is applied on the discovered top-K frequent attributes in the first phase, and ensures the differential privacy for entire notification events. Finally, the brokers utilize the top-K attributes (of each user) to match the appropriate publishers and subscribers. The proposed scheme can protect the privacy and confidentiality of users while maintaining the function of a typical PS system. We proved that the proposed scheme can ensure the differential privacy and resist the collusion attack. Moreover, we conduct the experiments on real world E-commerce datasets, and the results show that the PCP can efficiently achieves the trade-off between the system cost and the privacy demand.

In a nutshell, the main contributions in this paper are summarized as follows.

1) The PCP, a novel privacy-preserving PS scheme is proposed by using differential privacy in fog computing context, which can simultaneously ensure users' privacy, confidentiality and the function of publish–subscribe. Moreover, this scheme can resist the collusion attack and support uncertain datasets circumstance.
2) A comprehensive complexity analysis of the proposed scheme in terms of the data structure, differential privacy and publish–subscribe service framework is provided. The analysis results show that the proposed scheme is ϵ-differential private and can protect the system security. Meanwhile, this scheme can provide the publish–subscribe service stably.
3) To illustrate feasibility and availability, we conduct the experiments on real world datasets (E-commerce datasets in TIANCHI website [11]). The results demonstrate that the proposed scheme provides security with reasonable overhead. In other words, the system runtime overhead and the privacy demand can reach a comprehensible trade-off.

The remainder of this paper is organized as follows. In Section II, related works are reviewed. In Section III, the useful preliminaries are provided. The system model and design objectives are presented in Section IV. Section V elaborates the PCP. Section VI provides the security analysis. Performance evaluation based on real world datasets is provided in Section VII. Finally, concluding remarks are given in Section VIII.

II. RELATED WORK

In recent years, the PS system has been applied to many LSMCS as the key technology [2], [4]. The Conseil Europeen pour la Recherche Nucleaire (CERN) uses the operational grid activities (monitoring systems) of the large hadron collider (LHC) to integrate over 100,000 machines in 20 different countries so as to form a grid for processing operational monitoring data from the LHC and other scientific instruments of CERN [4]. The city of Tokyo utilizes highway traffic monitoring which interconnect roadway sensors and roadside kiosks to a centralized control center so as to deliver constant updates to kiosks and to gather traffic condition data from sensors [4]. The grand coulee dam establishes the power plant monitoring and control to interconnect 40,000 Supervisory Control And Data Acquisition (SCADA) systems controlling the 30 generators of the dam and the transmission switchyard [4].

In order to protect the the security of PS system, many security solutions have been designed [4], [5]. The intuitive method is through the encryption. Yang *et al.* proposed an attribute-keyword based access control scheme for data publish–subscribe in cloud [12]. Tariq *et al.* designed a broker-less PS system by using the identity-based encryption to ensure the security [23]. Tian *et al.* proposed a PS system composing of engine, subscription manager and matching engine to achieve security [14]. Nabeel *et al.* introduced a feasible solution to meet many constraints based on the public key cryptosystem [15].

In briefly, the aforementioned works only consider a specific scenario. With the emergence of internet of things (IoT), the security risks have drawn increasing attention [10], [24]–[27], [39]. To accommodate different circumstances (e.g., in the distributed environment) and higher

requirements (posed by the practical applications) to PS system, various countermeasures were proposed, such as [9] and [16]–[18]. Meanwhile, to solve the ever-growing security risks of PS system in new circumstances, a number of schemes were also proposed. Diro *et al.* proposed a lightweight scheme by using elliptic curve cryptography to ensure security in fog-based PS system [20]. A secure PS system that provides user data privacy by using hierarchical inner product encryption was proposed by Rajan *et al.* [21]. Beligianni *et al.* presented a solution that preserved consumer privacy in smart grids [22]. Due to the complex environment, more practical solutions need to be exploited. Moreover, the security threats (such as the collusion attack) still need to pay more attentions [13].

The differential privacy technology is a proper option to protect fog-based PS system security. Flourishing with the technology of big data and IoT, differential privacy becomes a hot area of research [19], [28]–[31]. Dwork and Roth discussed the differentially private methods for mechanism design and machine learning in [28]. Dwork reviewed the definition of differential privacy and provided a survey to the differential privacy frontier [29]. Zhang *et al.* proposed a differentially private method called PRIVBAYES for releasing high-dimensional data [30]. Li *et al.* presented an algorithm called PrivBasis that can find the most frequent itemsets with differential privacy [31].

Different from the aforementioned works, this work focuses on uncertain datasets of users, considers a PS system in fog-based context, adopts the differential privacy technique, achieves the privacy and security of users' data and prevents the collusion attacks in PS system.

III. PRELIMINARIES

This section reviews the main fundamental concepts related to our work, including frequent uncertain itemset mining [36] and differential privacy [33].

A. FREQUENT ITEMSET MINING

Frequent itemset mining (FIM) is utilized in data mining for exploring the frequent of itemsets in a given dataset. The FIM can discover all the greater than or equal to θ-frequency itemsets together with their frequencies by inputting θ parameter, or return the top K most frequent itemsets with their frequencies by inputting integer K. Most of the FIM algorithms for certain datasets are based on the *Apriori* algorithm [40]. Moreover, the *U-Apriori* algorithm, inherited from the *Apriori*, can be extended to deal with the uncertain datasets by using the concept of expected support count [36]. For convenience, let the abbreviation of U-FIM denotes the algorithm of frequent itemset mining for uncertain datasets in this paper.

Suppose that all the records in uncertain dataset are mutually independent, and let all uncertain items in the same record are mutually independent. For a set of possible database $D = \{d_1, d_2, \cdots, d_{|d|}\}$, each possible data $D_w (1 \leq w \leq |d|)$, Let $P(D_w)$ denote the probability of a possible data D_w. The $P(D_w)$ can be obtained by [36]:

$$P(D_w) = \prod_{i=1}^{n} \left(\prod_{x \in I(D_w,i)} P(x \in t_i) \cdot \prod_{y \notin I(D_w,i)} (1 - P(y \in t_i)) \right) \quad (1)$$

where $I(D_w, i)$ denotes the set of items that contained in record i and belonging to D_w.

Let the $S_e(X)$ denote the expected support of itemset X, it can be obtained by [36]:

$$S_e(X) = \sum_{i=1}^{|d|} P(D_i) \times S(X, D_i) \quad (2)$$

where $S(X, D_i)$ is the support counter of itemset X in possible data D_w.

B. DIFFERENTIAL PRIVACY

The differential privacy was designed to preserve the privacy of datasets. Namely, with the differential privacy mechanism, adding or removing a single item of datasets, the output of data analysis will be the same as input. Evidently, the output cannot be used by adversaries to gain access to users, data by using their background information [29].

Suppose two databases D_1 and D_2, are two neighboring databases that differ by no more than one record.

Definition 1 (ϵ-Differential Privacy [32]): For a randomized algorithm $\boldsymbol{A}$ gives ϵ-differential privacy if for any pair of neighboring datasets D_1 and D_2 of Hamming distance $d(D_1, D_2) \leq 1$, and any $S \in Range(\boldsymbol{A})$,

$$\Pr[\boldsymbol{A}(D_1) = S] \leq e^{\epsilon} \cdot \Pr[\boldsymbol{A}(D_2) = S] \quad (3)$$

where ϵ is the privacy budget of differential privacy.

The maximal possible difference value between the outputs of the pair of neighboring datasets can be obtained by using the sensitivity, defined as follows.

Definition 2 (Sensitivity [33]): Let $\boldsymbol{D}$ denote the space of all databases. For a given function $f : D \to R^d$, the sensitivity of f is:

$$\Delta f = \max_{D_1, D_2} \|f(D_1) - f(D_2)\|_1 \quad (4)$$

where D_1 and D_2 are any pair of neighboring datasets.

Lemma 1 (Composition Lemma [34]): For a given sequence of algorithm $f = f_1, f_2, \cdots, f_s$, if each algorithm $f_i (1 \leq i \leq s)$ can ensure ϵ_i-differential privacy, the algorithm f can ensure $\sum_{i=1}^{s} \epsilon_i$-differential privacy.

The approach to design the scheme that satisfy ϵ-differential privacy can generally be divided into the Laplace and exponential based mechanisms. Both of them will be used in this paper.

1) LAPLACE MECHANISM

The Laplace mechanism computes the function g on the dataset D and adds a random noise $Lap(\beta)$. The noise $Lap(\beta)$

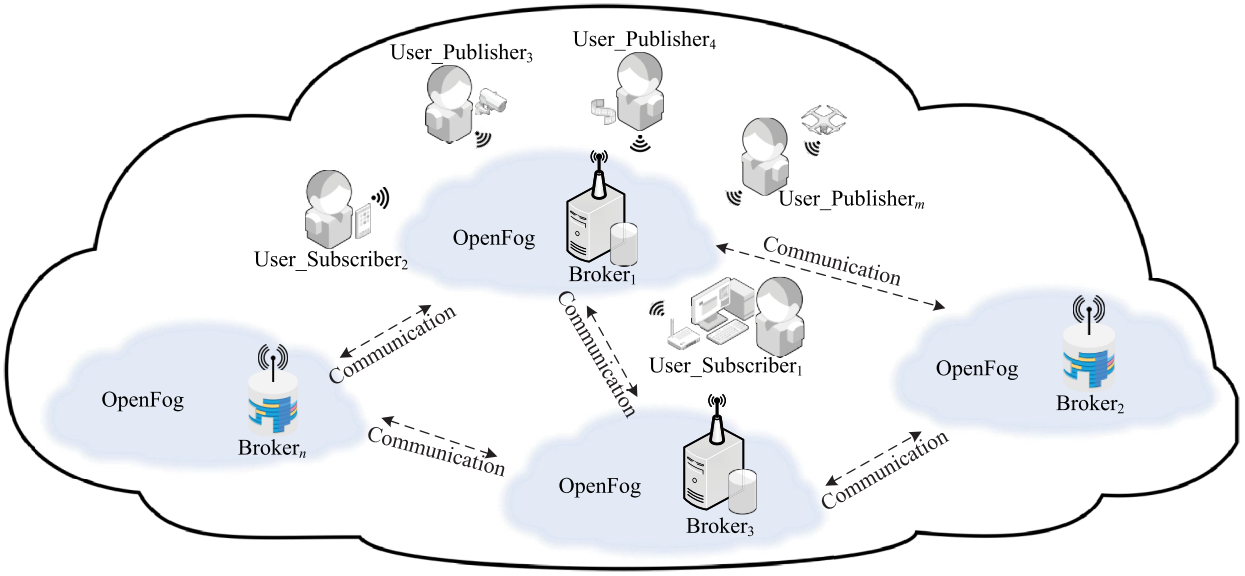


FIGURE 1. Network model.

denotes a random variable sampled from the Laplace distribution with scale parameter β. Its probability density function is $p(x) = \frac{1}{2\beta} \exp\ (-\frac{|x|}{\beta})$.

Definition 3 (Laplace mechanism [32]): For the given database D and a function $f : D \to R^d$, its sensitivity is Δf. The algorithm provides ϵ-differential privacy:

$$\boldsymbol{A}(D) = f(D) + Lap\left(\frac{\Delta f}{\epsilon}\right) \tag{5}$$

where $\Delta f = \max_{(D_1,D_2):D_1 \simeq D_2} |f(D_1) - f(D_2)|$, and $\Pr[Lap(\beta) = x] = \frac{1}{2\beta} \exp\ (\frac{-|x|}{\beta})$.

2) EXPONENTIAL MECHANISM

The exponential mechanism samples the set of all possible answers in the range of g according to an exponential distribution (the more accurate answers will be sampled with higher probability) and then computes a function g on a dataset D.

Definition 4 (Exponential Mechanism [35]): Given a randomized algorithm $\boldsymbol{M}$, input dataset D and output entity object $r \in Range$, for a quality function $q : D \times R \to \mathbb{R}$, its global sensitivity Δf_q is defined as $\Delta f_q = \max_r \max_{(D_1,D_2):D_1 \simeq D_2} \left|q(D, r) - q(D', r)\right|$. Then, the algorithm $\boldsymbol{M}$ satisfies ϵ-differential privacy:

$$\Pr[\boldsymbol{M}(D) = r] \propto \exp\left(\frac{\epsilon}{2\Delta f_q} q(D, r)\right) \tag{6}$$

where the real valued score $q(D, r)$ indicates how accurate it is to return r when the input dataset is D.

IV. SYSTEM MODEL AND DESIGN GOALS

In this section, we present the system model and the design goals.

A. NETWORK MODEL

We consider a fog-based PS scheme including N brokers and M users (including publishers and subscribers), as shown in Fig. 1. The fog-based PS system environment consists of users, brokers and many *OpenFogs*. Users communicate with each other through the *OpenFog* by connecting the broker in proximity. Content-based data dissemination is employed for event routing. Denote Λ be the event space which is composed of a set of t attributes B_t, where $\Lambda = \{B_1, B_2, \cdots, B_t\}$. Each attribute is characterized by an exclusive identifier and its data type (*i.e.*, integer, floating point, and character strings). An event is constituted by the different attributes and relevant values. A broker matches a given publisher and subscriber by comparing the attributes in their events.

1) *Broker* acts as the role of man-in-the-middle between the users in PS system. The broker receives the notification events of users and temporarily stores these events for sending to destination users at the appropriate time. Moreover, the broker matches these received notification events by analyzing the attributes and data, and then sends the matching notification to the matched users. Actually, the broker corresponds to the fog node in fog-based PS system circumstance. The power, computing and storage capacity at each fog node are limited.
2) *Users* include publishers providing the service or resource and subscribers consuming the service or resource in the local area. The publishers and subscribers are denoted by $\mathbb{P} = \{p_1, p_2, \cdots, p_M\}$ and $\mathbb{S} = \{s_1, s_2, \cdots, s_N\}$, respectively. Each user has itself profiles (e.g., unique identifier, attributes, notification events and processed datasets) by performing the system default initial procedure. Once receiving the matching notification from the broker, 1) the publisher

sends its resources (include its datasets or application interface) to broker, and 2) the subscriber sends its profile information to broker to confirm and connect the publish service.

B. SECURITY MODEL

Malicious users and brokers may involve in the PS system and launch attacks in the packet delivery and user matching. We define two types of attacks: 1) privacy leakage attack (PLA) and 2) collusion attack (CA). Some brokers may curiously collect and analyse users' preference and private profiles. PLA attempts to breach and reveal users' private and sensitive information. The privacy can be disclosed during matching the relative users, storage (caching the datasets) and the packet delivery phases. On the other hand, malicious users may collude with corrupted brokers to break the datasets of legitimate users. Accordingly, it is possible to deny using or providing service such that the malicious user may reject paying or deny of service. These two kinds of attacks, i.e., PLA and CA, can cause massive communication, computing and storage overheads. Meanwhile, it would destroy the confidentiality and feasibility of PS system.

1) ATTACKER MODEL

The attacker model is similar to the commonly used honest-but curious model [37] and the attacker has all the background information about users' datasets. We assume that the communication channels are non-secure channels, which is more practical. All the entities (i.e., publisher, subscribers, and brokers) are computationally bounded, and the brokers are considered to have more computing and storage capacities than the publishers and subscribers. Publishers and subscribers distrust each other but both trust the brokers. Nevertheless, brokers do not trust any publishers or subscribers. Moreover, all the entities are honest and perform functions following the designed protocol strictly. Additionally, normal brokers diffuse the proper events, while malicious brokers and users (the publishers or subscribers) may collude with each other.

1) *Broker colludes with publisher*: The malicious brokers and publishers may spread the fake or duplicate events to the overlay PS network. Moreover, malicious publisher can give other subscribers' confidential information to brokers (for the future data analysis), leading to the privacy leakage.
2) *Broker colludes with subscriber*: The malicious brokers and subscribers may deny admitting matching accomplishing or utilizing data and services from the legal publishers.

The published events in the PS scheme always be attractive to the inquisitive publishers. Likewise, subscribers are inquisitive to pry into the subscriptions of other subscribers and the events published by publishers which are not authorized to subscribe. Moreover, the corrupted broker may pry into the private information while its processing the disseminating and matching processes.

C. DESIGN GOALS

The design goal is to develop a fog-based privacy-preserving PS scheme. Specifically, we aim to achieve the following objectives:

1) *Practical Goals*: Confronted by the massive heterogeneous data and emerging network architecture with computing and storage resources at edge, our goal is to develop a practical scheme to ensure user data privacy in PS mechanism. The proposed scheme should be deployed appropriately in fog computing environment with minimal extra computational, storage and communication overheads to the users. Specifically, the privacy protection mechanism should not cause too much costs and maintenance can be performed efficiently and practically.
2) *Security Goals*: Our security goal is to preserve the user's privacy against PLA and CA. 1) The proposed scheme should protect the data privacy in dissemination. The adversary cannot discover private information by directly analyzing the transmitted event datasets. 2) The CA should not be able to forge the legitimate events or to deny the fact that it had been used the resources and services, because the data analysis output dose not change by adding or removing an item to or from the datasets.

V. PROPOSED PCP SCHEME

In this section, the PCP, a fog-based privacy-preserving PS scheme is proposed. This scheme can be divided into three main procedures, encompassing 1) notification event production, 2) data privacy protection, and 3) events matching.

As shown in Fig. 2, the framework of PCP includes the following steps.

1) *Notification Event Production*: The PS system generates the notification event for each user (each publisher and subscriber) respectively, which contains the users' content profiles (e.g., unique identifier and attributes). To prepare the notification event, the PS system employs the top-K U-FIM algorithm on user's uncertain datasets to mine the top-K most frequent itemsets. Then, the exponential mechanism of differential privacy technique is applied to these mined top-K most frequent itemsets for achieving differential privacy.
2) *Content-Based Event Privacy Protection*: Based on the mined frequent itemsets from uncertain datasets, the Laplace mechanism is applied to ensure differential privacy for the operated datasets.
3) *Events Matching*: The broker uses the attributes of the top-K most frequent itemsets to match the corresponding events.

For convenience, the key notations used in Section V are given in TABLE 1.

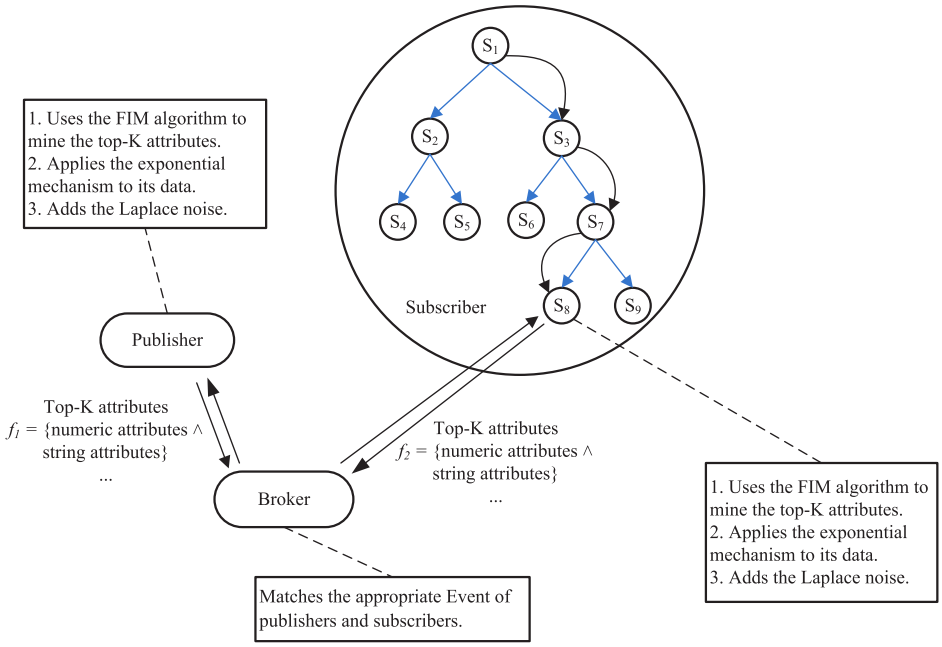


FIGURE 2. The framework of privacy-preserving content-based publish–subscribe scheme.

TABLE 1. Notations.

Symbol	Description
S	a set of uncertain dataset
n	the size of dataset
m	the size of alphabet
l	the length of itemset
μ	the error parameter
δ	the confidence parameter
K	the number of specified patterns
f_K	the expected support of the K_{th} most frequent itemset
$S_e(\tau)$	the expected support of itemset τ
$\tilde{S}_e(\tau)$	the noise expected support of itemset τ
ϵ	the privacy budget for differential privacy
$\tilde{f}(\tau)$	the truncated expected support of itemset τ

A. NOTIFICATION EVENT PRODUCTION

In notification event production procedure, PS system generates the notification event, by means of U-FIM and differential privacy, to each publisher and subscriber respectively. The notification event includes the attribute list of top-K most frequent itemsets, unique identifier of user and the timestamp. The proposed notification event production approach, mines the top-K most frequent itemsets and protects these frequent itemsets in the uncertain datasets. Specifically, it includes three steps as follows.

1) The PS system runs the U-FIM algorithm for uncertain datasets to get the K itemsets from datasets with the expected support greater than or equal to $f_K = \psi$. To mine the desired K itemsets, this approach utilizes the *U-FIM* algorithm twice. The first round applying U-FIM is to get the frequent threshold f_K.
2) The second round is to get all the itemsets with the expected support greater than or equal to $f_K - \mu$. According to Lemma 3 (explained in Section VI), the error parameter $\mu = \frac{4K}{\epsilon}\left(\ln\frac{K}{\delta} + \ln\binom{m}{l}\right)$. For every itemset mined by PS system, it has to go through the entire dataset once in this step.
3) After determining the top-K most frequent itemsets, the PS system samples these itemsets from the uncertain dataset without replacement. In this step, we use the truncated expected support to sample K itemsets. According to Definition 4 and Equation (6), for the probability of choosing a specific itemset denoted as $\Pr(\delta)$, it holds that $\Pr(\delta) \propto \exp\left(\frac{\epsilon}{4K}\tilde{f}(\delta)\right)$.
4) For the notification event of a given user, PS system adds the rest information (e.g., the unique identifier of user and timestamp) to it.

Algorithm 1 illustrates the notification event production procedure in the privacy preserving PS scheme.

Algorithm 1 Producing Notifications

Input: Uncertain dataset S, set of items τ, datasets size N, top K, privacy budget ϵ, itemset length l, expected support of the K_{th} most frequent itemset f_K, truncated expected support $\tilde{f}$, and error parameter μ

1: **Preprocessing:** Initialization and applying U-FIM algorithm to find all the itemsets from uncertain datasets with the expected support $> f_K - \mu$.
2: **Sampling and adding noise:** Sample the K itemsets with $\Pr[\tau] \propto exp\left(\frac{\epsilon}{4K}\tilde{f}(\tau)\right)$ and without replacement. Add $Lap\left(\frac{2K}{\epsilon}\right)$ noise to the sampled itemsets.
3: **Packaging:** Package the notification event with adding the unique identifier of user and timestamp.

Algorithm 2 Applying U-FIM

procedure ApplyUFIMAlgorithm($S, \epsilon, l, K, f_K, \mu$)
  function GetExpectedSupport($S, f_K, \tau, \tilde{f}, N$)
    $K_{th} \leftarrow \emptyset$;
    for each $\tau \in S$ **do**
      $\tilde{f}(\tau) \leftarrow ESupp(\tau)$;
      if $\tilde{f}(\tau) \geq f_K$ and $|N \cup K_{th}| < K$ **then**
        $K_{th} \leftarrow K_{th} \cup \{\tau\}$;
      end if
    end for
    return K_{th};
  end function

  function FindFrequentItemsets($S, n, \epsilon, f_K, \tau$)
    $L_1 \leftarrow$ GetExpectedSupport (S, f_K, τ);
    $l \leftarrow 2, L \leftarrow L_1$;
    while $L_{l-1} \neq \emptyset$ and $|L| < K$ **do**
      $S_l \leftarrow \{\alpha \cup \beta | \alpha, \beta \in L_{i-1} \wedge \alpha < \beta\}$;
      $L_l \leftarrow$ GetExpectedSupport (S_l, f_K, τ, L);
      $L \leftarrow L \cup L_l, l \leftarrow l + 1$;
    end while
    return L;
  end function
end procedure

Namely, Algorithm 2 demonstrates the procedure of applying the U-FIM algorithm.

Algorithm 3 demonstrates the procedure of applying exponential mechanism.

B. CONTENT-BASED EVENT PRIVACY PROTECTION

After generating the notification event, the top K most frequent itemsets and their frequencies are discovered. In order to protect the differential privacy of datasets, the scheme applies the Laplace mechanism of differential privacy technique which utilizes a zero mean Laplace noise with the parameter $\frac{2K}{\epsilon}$ to disturb the true frequencies of the top K sampled itemsets.

According to Definition 4, Lemma 2 and Equation (5), we can obtain

$$\begin{aligned} \check{S}_e(\tau_i) &= S_e(\tau_i) + Lap\left(\frac{K}{\epsilon/2}\right) \\ &= S_e(\tau_i) + Lap\left(\frac{2K}{\epsilon}\right) \end{aligned} \tag{7}$$

where $\tau_i\,(1 \leq i \leq K)$ is the itemset of discovered top K frequent itemsets from uncertain datasets, and $T = (\tau_1, \tau_2, \ldots, \tau_K)$.

Algorithm 4 illustrates the details of applying Laplace mechanism on event data in the privacy preserving PS scheme.

C. EVENT MATCHING

In the following, we elaborate the matching approach as follow. Since the attributes of each event has already been discovered, the matching approach uses these attributes to match the correlative events by comparing them. Namely, the attributes are prepared to match the top K sampled itemsets. These attributes can be divided into three categories, including numeric, string and complex attributes [23]. The details of these three types matching will be described in the following.

Algorithm 3 Applying Exponential Mechanism

procedure Sampling and adding($S_{>f_k-\mu}, \epsilon, l, K, f_K, \mu, N$)
  $N \leftarrow \left|S_{>F_k-\mu}\right| + 1$;
  Create an array $A\,[1, \cdots, N-1]$;
  for $i = 1$ to $N-1$ **do**
    $A_i.itemset \leftarrow S_{>F_k-\mu}\,(i)$;
    $A_i.ESupp \leftarrow ESupp\left(S_{>F_k-\mu}\,(i)\right)$;
    $A_i.expData \leftarrow exp\left(\frac{\epsilon \cdot ESupp\left(S_{>F_k-\mu}(i)\right)}{4K}\right)$;
  end for
  $A_N.itemset \leftarrow lowESuppItems$;
  $\left(\binom{m}{l} - \left|S_{>F_k-\mu}\right|\right) exp\left(\frac{\epsilon \cdot (f_K-\mu)}{2K}\right)$;
  Create a doubly linked list DbL with N nodes and stores DbL_i and $\sum_{i \leq j \leq N} A_j.expData$ in it;
  $Banned \leftarrow \emptyset$, $Output \leftarrow \emptyset$;
  for $i = 1$ to K **do**
    $flag \leftarrow TURE, j \leftarrow 1$;
    while $flag \leftarrow TURE$ **do**
      Generate $Y \sim Bernoulli\left(\frac{A_j.expData}{X_j}\right)$;
      if $N == j$ **then**
        $flag \leftarrow FALSE$;
        Randomly sample itemset τ from the collection of all length l itemsets;
        $Banned \leftarrow Banned \cup \tau$, $Output.itemset \leftarrow \tau$, $Output.ESupp \leftarrow f - K - \mu$;
        Updata A_N and X_q;
      else if Y == 1 **then**
        $Output.itemset \leftarrow A_j.itemset$, $Output.ESupp \leftarrow A_j.ESupp$;
        Updata X_q, Remove Node DbL_j and $N \leftarrow N-1$;
        $flag \leftarrow FALSE$;
      end if
      $j \leftarrow j + 1$;
    end while
  end for
  for $t = 1$ to $N-1$ **do**
    $Output.ESupp \leftarrow A_j.ESupp + Lap\left(\frac{2K}{\epsilon}\right)$;
    $t \leftarrow t + 1$;
  end for
  return $Output$;
end procedure

1) NUMERIC ATTRIBUTES

For the γ different numeric attributes in a given event, the γ-dimensional spatial indexing approach [23] can be used to

Algorithm 4 Applying Laplace

Input: Uncertain dataset S, set of items τ, datasets size N, top K, privacy budget ϵ, itemset length l, expected support of the K_{th} most frequent itemset f_K, truncated expected support $\tilde{f}$, and error parameter μ

1: **procedure** ApplyLaplace($S_{>f_k-\mu}, \epsilon, l, K, f_K, \mu, N$)
2: $N \leftarrow \left|S_{>f_k-\mu}\right|, X \leftarrow \emptyset, \sigma \leftarrow f_k - \mu$;
3: **return** $P_{score}\,(c)_{max}$;
4: **for** $i = 1$ to N **do**
5: $X_i.itemset \leftarrow S_{f_K-\mu}\,(i)$, $X_i.ESupp \leftarrow ESupp\,\left(S_{f_K-\mu}\,(i)\right)$;
6: **end for**
7: $l_{ESupp} \leftarrow K$-th highest noisy expected support in X;
8: **if** $l_{ESupp} \geq \sigma$ **then**
9: $p \leftarrow \frac{1}{2}exp\left(-\frac{|\sigma - l_{ESupp}|\epsilon}{4K}\right)$;
10: **else**
11: $p \leftarrow 1 - \frac{1}{2}exp\left(-\frac{|\sigma - l_{ESupp}|\epsilon}{4K}\right)$;
12: **end if**
13: $Y \sim Binom\left(\binom{m}{l} - N, p\right), Banned \leftarrow \emptyset$;
14: **for** $i = N + 1$ to $N + 1 + Y$ **do**
15: Randomly sample itemset τ from the collection of all length l itemsets;
16: $Banned \leftarrow Banned \cup \tau$, $X_i.itemset \leftarrow \tau$, $X_i.ESupp \leftarrow \sigma$;
17: $X_i.noisyESupp \sim$ exponential distribution with mean $l_{ESupp} + \frac{4K}{\epsilon}$;
18: **end for**
19: Set the top-K itemsets to $Output$;
20: **return** $Output$;
21: **end procedure**

process such numeric attributes matching. The spatial indexing approach classifies the event space into regular subspaces, which makes the space as an enclosed approximation for the subscriptions and advertisements, as shown in Fig. 3.

The subspaces are labeled by a bit string of "0" and "1". A planar subspace Z_1 covers the planar subspace Z_2, where Z_2 is a suffix of Z_1. For a subscription or advertisement, it can be represented by one or more subspaces composed of peer. For instance, as shown in figure 3, Sp_1 is mapped to one subspace with the label {11} and Sp_2 is mapped to two subspaces with the labels {110, 111}. Moreover, for a given event, it probably has large amounts of numeric attributes. Accordingly, the number of subspaces that these numeric attributes mapped to would be very large. In order to solve this issue, PCP decomposes the domain of each attribute into subspaces separately. The binary tree can be built separately for each attribute to solve the attribute decomposition, as shown in Fig. 3.

2) STRING ATTRIBUTES

The above introduced spatial indexing approach can also deal with the data type of string when the data type has a known domain. Fortunately, string attributes usually have the statistics of its maximum number of characters and this feature

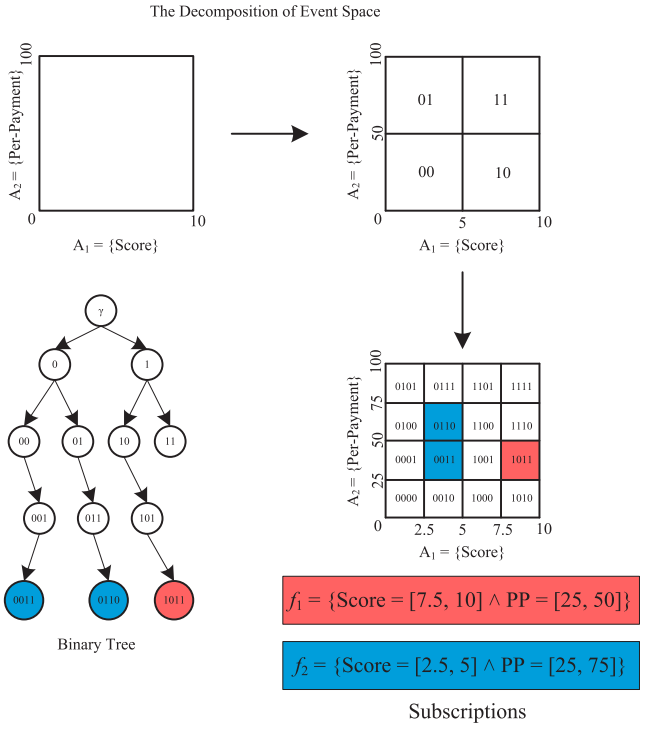


FIGURE 3. Numeric attributes.

makes the string attributes to have known bounds. By using the hashing technique or other linearization methods [23], the string attributes can be linearized and consequently be indexed. For the sake of indexing and matching the attribute of the event, the prefix matching dictionary tree has been built, as shown in Fig. 4. Each node in this tree is appointed to one single character string, which is same as its subscription or advertisement. All the events correspond to the leaf nodes of the prefix matching dictionary tree. In brief remarks, the longest path from leaf node to the root of a given attribute A_i in this dictionary tree is L_i, where L_i is the length of the longest string appointed to a leaf node. Intuitively, the suffix matching can similarly apply this approach.

3) COMPLEX ATTRIBUTES

In practical applications, the event always contains both type of attributes, i.e., the numeric and string, and the PCP defines a method of conjunction on these different predicates. A subscription or advertisement matches an event if and only if all the attributes successfully satisfy its predicates. For instance, consider an advertisement $f_1 = \{Score = [4, 5] \wedge Categoryname = Snack\}$ and a subscription $f_2 = \{PP = [80, 100] \wedge Cityname = Chengdu\}$, where PP is $Per\text{-}Payment$. The proposed approach 1) finds the numeric and string attribute respectively, and then 2) evaluates the conjunction predicates whether the conditions are true or not. The proposed approach successfully matches an event and an advertisement or subscription if and only if there exists an event satisfying the conjunction predicates.

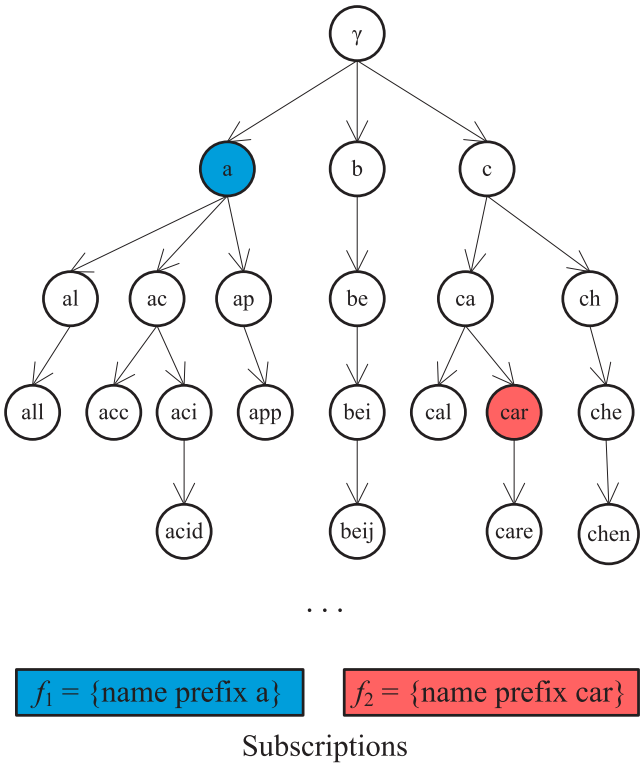


FIGURE 4. The example of prefix matching dictionary tree.

VI. SECURITY ANALYSIS

This section discusses the security properties and proves that the proposed privacy-preserving PS scheme is ϵ-*differentially private*.

We obtain the sensitivity of truncated expected support of itemset in Lemma 2 as follows.

Lemma 2: Let T and $\tilde{T}$ be the two of n uncertain transaction datasets with only one different record. Let $f^T(\tau)$ and $f^{\tilde{T}}(\tau)$ represent the expected support of an itemset τ in T and $\tilde{T}$, respectively. Let f_K^T and $f_K^{\tilde{T}}$ is the expected support of the K-th most frequent itemsets for uncertain dataset T and $\tilde{T}$, respectively. $I = \left\{\tau | \tau \in T \cap \tau \in \tilde{T}\right\}$ be the intersection of T and $\tilde{T}$. m is the record number of uncertain datasets and t_j $(1 \le j \le m)$ represents a record in uncertain datasets. Then, we state that the sensitivity of truncated expected support of any itemset τ is 1.

Proof: According to Equation (2) in Section III, we can calculate the expected support of itemset τ [36]:

$$f^T(\tau) = \sum_{j=1}^{m} \prod_{t \in T} P\left(\tau \in t_j\right) \tag{8}$$

For $I_1 = \left\{\tau | \tau \notin T \cap \tau \in \tilde{T}\right\}$ and $I_2 = \left\{\tau | \tau \in T \cap \tau \notin \tilde{T}\right\}$, we have

$$\begin{cases} f_K^T = f_K^T + \alpha_1 \\ f_K^{\tilde{T}} = f_K^{\tilde{T}} + \alpha_2 \end{cases} \Rightarrow \left\| f_K^T - f_K^{\tilde{T}} \right\|_1 = \|\alpha_1 - \alpha_2\|_1 \le 1 \tag{9}$$

According to Equation (8), we can obtain the expected support of itemset τ in uncertain datasets T and $\tilde{T}$, respectively:

$$\begin{aligned} f^T(\tau) &= \sum_{j=1}^{|T|} \prod_{t \in T} P\left(\tau \in t_j\right) \\ &= \sum_{j=1}^{|I|} \prod_{t \in T} P\left(\tau \in t_j\right) + \prod_{t \in T} P\left(\tau \in I_1\right) \end{aligned} \tag{10}$$

$$\begin{aligned} f^{\tilde{T}}(\tau) &= \sum_{j=1}^{\left|\tilde{T}\right|} \prod_{t \in T} P\left(\tau \in t_j\right) \\ &= \sum_{j=1}^{|I|} \prod_{t \in T} P\left(\tau \in t_j\right) + \prod_{t \in T} P\left(\tau \in I_2\right) \end{aligned} \tag{11}$$

As inequalities $f^T(\tau) \ge f_K^T - \beta$ and $f^{\tilde{T}}(\tau) \ge f_K^{\tilde{T}} - \beta$ hold, we have

$$\left\| f^T(\tau) - f^{\tilde{T}}(\tau) \right\|_1 = \left\| \prod_{t \in T} P(\tau \in I_1) - \prod_{t \in T} P(\tau \in I_2) \right\| \le 1 \tag{12}$$

As inequalities $f^T(\tau) \ge f_K^T - \beta$ and $f^{\tilde{T}}(\tau) < f_K^{\tilde{T}} - \beta$ hold, we have

$$\begin{aligned} f^T(\tau) - f^{\tilde{T}}(\tau) &= f^T(\tau) - f^{\tilde{T}}(\tau) + \beta \\ &\le f^T(\tau) - \beta + f^{\tilde{T}}(\tau) + \beta \\ &\le 1 \end{aligned} \tag{13}$$

and

$$f^T(\tau) - f^{\tilde{T}}(\tau) \ge -1 \tag{14}$$

From Equation (13) and Equation (14), we can obtain

$$\left\| f^T(\tau) - f^{\tilde{T}}(\tau) \right\|_1 \le 1 \tag{15}$$

Since $f^T(\tau) < f_K^T - \beta$ and $f^{\tilde{T}}(\tau) < f_K^{\tilde{T}} - \beta$ hold, we have

$$\left\| f^T(\tau) - f^{\tilde{T}}(\tau) \right\|_1 = \left\| f^T(\tau) - \beta - f^{\tilde{T}}(\tau) + \beta \right\|_1 \le 1 \tag{16}$$

To sum up, the sensitivity of truncated expected support of any itemset τ is 1. ■

The following Lemma 3 defines the error parameter μ used in PCP.

Lemma 3: Let T be the set of sampled itemsets in U-FIM sampling procedure and l be the maximal length of the itemsets in T. For all $\delta > 0$, with probability at least $1 - \delta$, the expected support of all the itemsets in T are greater than $f_K - \mu$, where $\mu = \frac{4K}{\epsilon}\left(\ln\frac{K}{\delta} + \ln\binom{m}{l}\right)$.

Proof: Suppose that the itemset, with the truncated expected support g, has not been sampled. Then the probability of an itemset has been sampled is less than $exp\left(-\frac{\epsilon\mu}{4K}\right)$ [34], where the truncated expected support of this itemset is less than $g - \delta$. In the whole sampling procedure,

the maximum probability of sampling the itemset that its truncated expected support is less than $g-\delta$ is $n^l \cdot exp\left(-\frac{\epsilon\mu}{4K}\right)$ [34]. This is because the numbers of itemset with the truncated expected support $< g-\delta$ are no more than n^l. For all the K sampled itemsets, the maximum probability of these itemsets with the expected support $\leq f_K - \mu$ is $K \cdot n^l \cdot exp\left(-\frac{\epsilon\mu}{4K}\right)$. Let $\delta \geq K \cdot n^l \cdot exp\left(-\frac{\epsilon\mu}{4K}\right)$, and then we can obtain $\mu \geq \frac{4K}{\epsilon} \cdot \left(\ln\frac{K}{\delta} + \ln\binom{m}{l}\right)$. ■

Theorem 1: Algorithm 1 is $\frac{\epsilon}{2}$-differentially private.

Proof: The third step in Algorithm 1, i.e., sampling the K itemsets without replacement, successively performs the exponential mechanism (explained in section III) for K times essentially. According to Lemma 2, the sensitivity of the truncated expected support of any itemset τ is 1. Thus, the sensitivity of the truncated expected support of top-K frequent itemsets is K. Moreover, according to Definition 4, the probability of the no-replacement sampled itemsets meet the condition of $\Pr[M(T) = \tau] \propto exp\left(\frac{(\epsilon/2)\cdot\tilde{f}(\tau)}{2K}\right) = exp\left(\frac{\epsilon\cdot\tilde{f}(\tau)}{4K}\right)$. As a result, the Algorithm 1 is $\frac{\epsilon}{2}$-differentially private. ■

Theorem 2: Algorithm 2 is $\frac{\epsilon}{2}$-differentially private.

Proof: As the proof of Theorem 1 that the sensitivity of the truncated expected support of top-K frequent itemsets is K. According to Definition 3, the scale parameter is $\frac{2K}{\epsilon}$. Hence, algorithm 2 is $\frac{\epsilon}{2}$-differentially private. ■

Theorem 3: The proposed privacy preserving content-based publish/subscribe scheme is ϵ-differentially private.

Proof: According to Theorem 1, Theorem 2 and Lemma 1, we can obtain that the proposed privacy preserving content-based publish/subscribe scheme is ϵ-differentially private. ■

VII. PERFORMANCE EVALUATION

We conduct the experiments to evaluate the performance of the proposed privacy preserving content-based publish/subscribe scheme.

A. SIMULATION SETUP

We evaluate the performance of PCP using a desktop computer with a 3.30 GHz Intel CPU, 16GB RAM, and windows 7 OS. A python environment is built to simulate the U-FIM, differential privacy and content-based publish–subscribe (CBPS) system. We utilize two reak-world datasets to simulate the communications of fog-based CBPS system and verify the availability and efficiency while protecting the data privacy. These datasets can be obtained from the TIANCHI website [11]. The parameters of these datasets are given in table 2, where the number of items is denoted as m and the number of operations on dataset is denoted as n. In order to add uncertainty to some of these datasets, an existential random probability in the range of (0, 1], is assigned to each item in each operation.

B. SIMULATION RESULTS

We first measure the average time delay of the subscribers to connect to fog-based PS system. The time delay is measured

TABLE 2. DataSets.

Field	m	n	Description
shop_id	2305	46535	the ID of seller
city_name	236	6852	the city name
location_id	1984	35489	the ID of sellers' location
per_pay	6544	842120	consumption per person
score	1574	55742	evaluation score
comment_cnt	842	8721	the quantity of evaluation
shop_level	10	76231	the level of the shop
cate_1_name	6214	125486	the name of category 1
cate_2_name	9574	6734	the name of category 2
cate_3_name	468	4116	the name of category 3
user_id	100255	51259	the ID of user
time_stamp	452	56325	time of payment
shop_id_paid	821	23684	the shop ID of user paid
location_id_user	6751	3236	the location ID of user

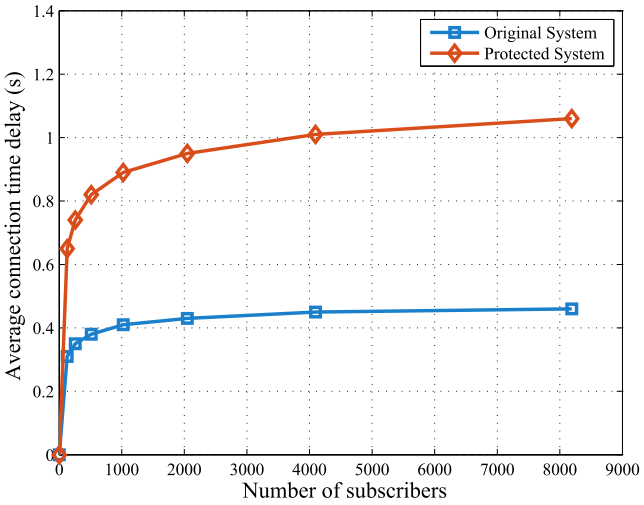


FIGURE 5. The average connection time delay of the subscribers.

from the time that a subscriber begins to connect to an appropriate broker in an *OpenFog* till the time that the subscriber and broker are connected successfully. Fig. 5 shows that the average connection time delay increases with the number of subscribers in the PS system, because more subscribers lead to more communication and computation overhead (e.g., increasing the breadth and depth of attribute tree). As shown in Fig. 5, in the beginning, the average connection time delay of the original PS system increases rapidly. Then, it rises along with the increasing of number of subscribers very slightly. Similarly, the average connection time delay of protected system (the proposed PCP) first increases rapidly and then the growth becomes slow. This is because the PS system can generate each attributes binary tree in parallel and each subscriber can connect the system simultaneously.

Then, we evaluate the average event propagation time delay of subscribers. The time delay of each subscriber is measured from the propagation of event by the publisher till the subscriber is successfully matched and served. As shown in Fig. 6, the average event propagation time delay of the original and protected system increases with the number of subscriber, and the protected system costs more time than

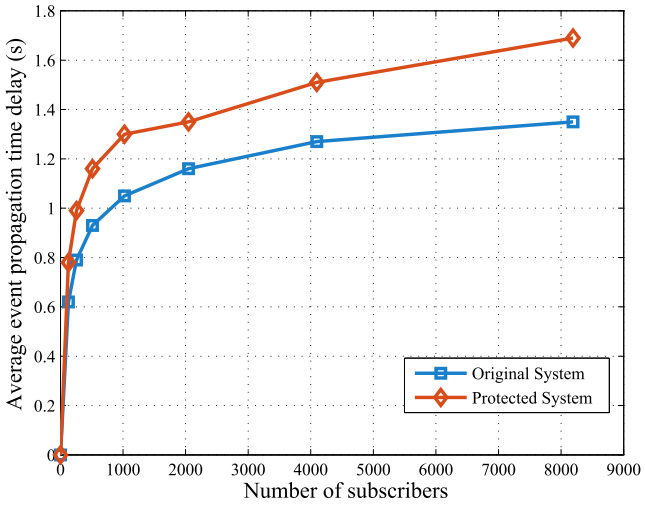


FIGURE 6. The average event propagation time delay of the subscribers.

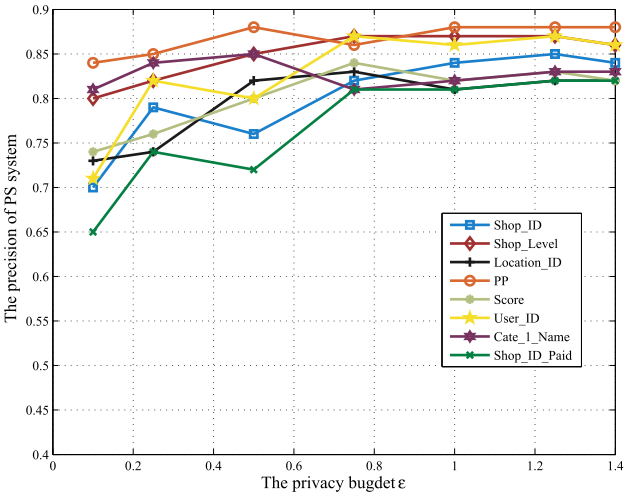


FIGURE 7. The precision by varying privacy budget ϵ with $K = 6$.

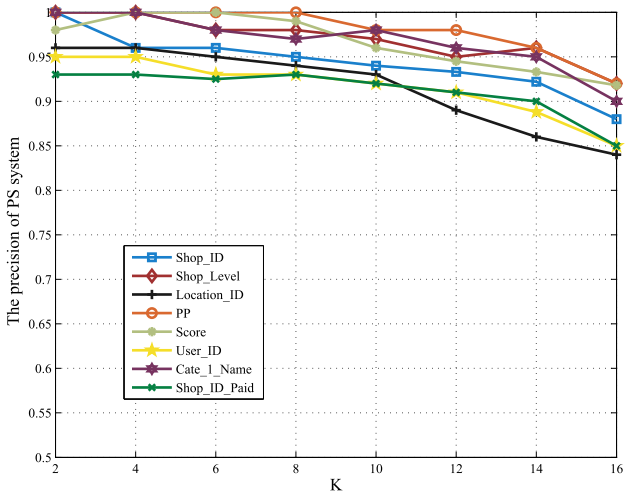


FIGURE 8. The precision by varying K with privacy budget $\epsilon = 0.8$.

the original system intuitively. This is because the increasing number of relevant subscribers is the same as the increasing depth of the propagation tree.

Next, we validate the precision of PCP with different K. The precision is used to assess the accuracy of mined top-K attributes of user (publisher or subscriber), which is defined by Equation (17) as follows [19]:

$$Precision = \frac{|U \cap U'|}{|U'|} \tag{17}$$

where U is the set of top-K frequent attributes and U' is the set of the frequent attributes obtained by PCP. Fig. 7 shows that, for $K = 6$, the precision of all attributes (itemsets) fluctuates, which is within the scope of (0.82, 0.96) with the increase of the privacy budget ϵ from 0.75 to 1.4. According to Definition 3 and Equation (5), a larger value of the privacy budget ϵ means a smaller Laplace noise, leading to more precise results. Moreover, according to Definition 4 and Equation (6), the probability of the top-K frequent itemsets has been chosen increases with the privacy budget ϵ, and a larger probability corresponds to a greater level of precision. However, as shown in Fig. 8, it can be seen that a larger value of K means a lower precision. The precision can maintain stability at the beginning, however, it begins to decline with the increasing of K. This is because, according to Equation (2), a larger value of K leads to a greater probability of low expected support frequent itemset has been chosen. As a result, the precision drops.

After that, we compare PCP with the MESA approach, in terms of the average matching time [38]. The MESA approach is used to contrast the effect of matching as it focuses on matching of both certain datasets and uncertain datasets in the PS system. Note that, most methods are focusing on the certain datasets. The matching time delay is measured from the time that a subscriber sends a subscribe request to the system till the time that a broker successfully matches the subscriber and publisher. Fig. 9 shows that the average matching time delay of all the three methods increase with the number of subscribers. The two matching time curves for the original system and MESA method are very close. The matching time of the protected system obviously exceeds the other two methods, and the reason for this result is that there is an overhead of approximately 0.16-0.23 seconds due to privacy mechanism.

Finally, we compare the PCP with the identity-based PS system (short as *IBE system*) [23], in terms of the overall average time delay. *IBE system* is designed as a broker-less mechanism to prevent the man-in-the-middle attack. However, this mechanism cannot address the shortage of computational and storage capability for the users. We measure the average time delay of the end-user equipments denoted as *User Equipment Only*, As shown in Fig. 10, the time delay of the *protected system* and the *User Equipment Only* increase fiercely with the number of subscribers. Nevertheless, the time delay of the *original system* and the *User Equipment Only* increase very slightly with the number of

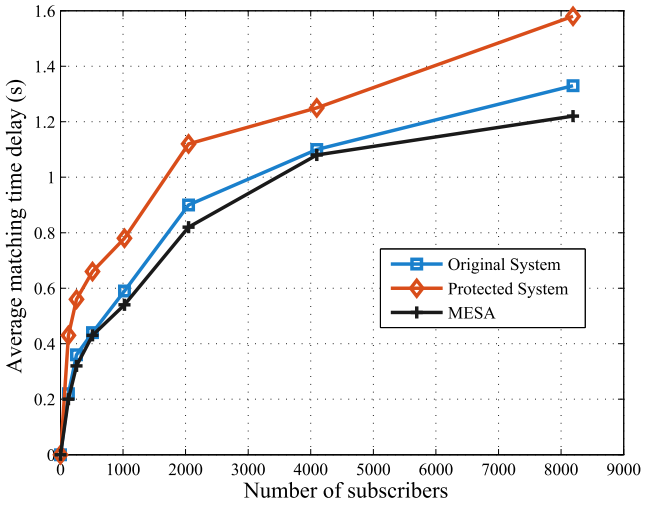


FIGURE 9. The average matching time delay of system.

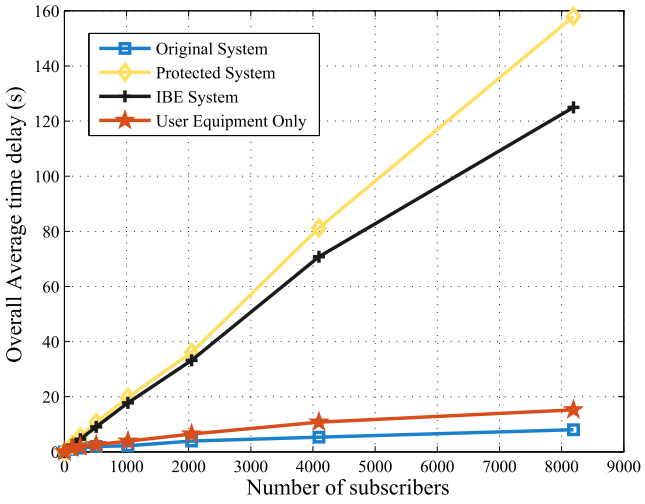


FIGURE 10. The overall average time delay of PS systems.

subscribers. The reason is that both the *protected system* and the *IBE system* utilize the algorithms that require lots of computing and storage capacity (e.g., the differential privacy and identity-based encryption algorithms). Fortunately, thanks to the fog computing, PCP can transfer the computing and storage overheads to the brokers (i.e., the fog nodes), which can greatly improve the feasibility and availability of PCP in practical applications.

VIII. CONCLUSION

In this paper, a privacy preserving content-based publish/subscribe scheme has been proposed to achieve the security and privacy protection in fog computing context. The proposed scheme 1) finds the top-K attributes of each event by utilizing U-FIM, 2) achieves the differential privacy by utilizing exponential and Laplace mechanisms, and 3) utilizes the complex attributes matching method to match the event of users appropriately. Security analysis demonstrated that the proposed PCP scheme can ensure the differential privacy and security. Finally, the experiments have been conducted to evaluate the performance of PCP. The results showed that the proposed privacy preserving CBPS scheme can achieve the privacy protection and alleviate user cost of computing and storage. These features make the proposed PCP scheme feasible and available in fog computing applications.

REFERENCES

[1] P. Bellavista, A. Corradi, and A. Reale, "Quality of service in wide scale publish—Subscribe systems," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1591–1616, 3rd Quart., 2014.

[2] E. Onica, P. Felber, H. Mercier, and E. Riviére, "Confidentiality-preserving publish/subscribe: A survey," *ACM Comput. Surv.*, vol. 49, no. 2, p. 27, 2016.

[3] I. Stojmenovic and S. Wen, "The fog computing paradigm: Scenarios and security issues," in *Proc. Federated Conf. Comput. Sci. Inf. Syst. (FedCSIS)*, Sep. 2014, pp. 1–8.

[4] C. Esposito and M. Ciampi, "On security in publish/subscribe services: A survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 966–997, 2nd Quart., 2015.

[5] A. V. Uzunov, "A survey of security solutions for distributed publish/subscribe systems," *Comput. Secur.*, vol. 61, pp. 94–129, Aug. 2016.

[6] A. Friedman and A. Schuster, "Data mining with differential privacy," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2010, pp. 493–502.

[7] C. Dwork and J. Lei, "Differential privacy and robust statistics," in *Proc. 41st Annu. ACM Symp. Theory Comput.*, 2009, pp. 371–380.

[8] P. Kairouz, S. Oh, and P. Viswanath, "The composition theorem for differential privacy," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 4037–4049, Jun. 2017.

[9] R. Lu, K. Heung, A. H. Lashkari, and A. A. Ghorbani, "A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT," *IEEE Access*, vol. 5, pp. 3302–3312, 2017.

[10] K. Yang, Q. Han, H. Li, K. Zheng, Z. Su, and X. Shen"An efficient and fine-grained big data access control scheme with privacy-preserving policy," *IEEE Internet Things J.*, vol. 4, no. 2, pp. 563–571, Apr. 2017.

[11] Alibaba Group. (Mar. 2017). *Ali Mobile Rec*. [Online]. Available: https://tianchi.aliyun.com/datalab/index.htm

[12] K. Yang, K. Zhang, X. Jia, M. A. Hasan, and X. S. Shen, "Privacy-preserving attribute-keyword based data publish-subscribe service on cloud platforms," *Inf. Sci.*, vol. 387, pp. 116–131, May 2017.

[13] R. Lu, H. Zhu, X. Liu, J. K. Liu, and J. Shao, "Toward efficient and privacy-preserving computing in big data era," *IEEE Netw.*, vol. 28, no. 4, pp. 46–50, Jul./Aug. 2014.

[14] Y. Tian, B. Song, M. M. Hassan, and E.-N. Huh, "An efficient privacy preserving pub-sub system for ubiquitous computing," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 12, no. 1, pp. 23–33, 2013.

[15] M. Nabeel, S. Appel, E. Bertino, and A. Buchmann, "Privacy preserving context aware publish subscribe systems," in *Proc. Int. Conf. Netw. Syst. Secur.*, 2013, pp. 465–478.

[16] A. Hakiri, P. Berthou, A. Gokhale, and S. Abdellatif, "Publish/subscribe-enabled software defined networking for efficient and scalable IoT communications," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 48–54, Sep. 2015.

[17] A. Antonić, M. Marjanović, K. Pripužić, and I. P. Žarko, "A mobile crowd sensing ecosystem enabled by CUPUS: Cloud-based publish/subscribe middleware for the Internet of Things," *Future Generat. Comput. Syst.*, vol. 56, pp. 607–622, Mar. 2016.

[18] V. N. Pham and E. N. Huh, "A fog/cloud based data delivery model for publish-subscribe systems," in *Proc. Int. Conf. Inf. Netw. (ICOIN)*, Jan. 2017, pp. 477–479.

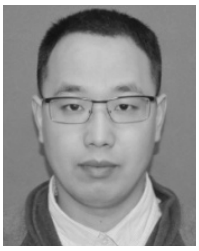
[19] Z. Ding, Z. Qin, and Z. Qin, "Frequent symptom sets identification from uncertain medical data in differentially private way," *Sci. Program.*, vol. 2017, May 2017, Art. no. 7545347.

[20] A. A. Diro, N. Chilamkurti, and N. Kumar, "Lightweight cybersecurity schemes using elliptic curve cryptography in publish-subscribe fog computing," in *Mobile Networks and Applications*. New York, NY USA: Springer, 2017, pp. 1–11.

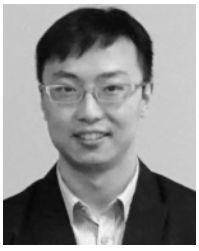
[21] M. A. Rajan *et al.*, "Security and privacy for real time video streaming using hierarchical inner product encryption based publish-subscribe architecture," in *Proc. 30th Int. Conf. Adv. Inf. Netw. Appl. Workshops (WAINA)*, 2016, pp. 373–380.

[22] F. Beligianni, M. Alamaniotis, A. Fevgas, P. Tsompanopoulou, P. Bozanis, and L. H. Tsoukalas, "An Internet of Things architecture for preserving privacy of energy consumption," in *Proc. Medit. Conf. Power Generat., Transmiss., Distrib. Energy Convers. (MedPower)*, 2016, p. 107.
[23] M. A. Tariq, B. Koldehofe, and K. Rothermel, "Securing broker-less publish/subscribe systems using identity-based encryption," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 2, pp. 518–528, Feb. 2014.
[24] Q. Wang, D. Chen, N. Zhang, Z. Qin, and Z. Qin, "LACS: A lightweight label-based access control scheme in IoT-based 5G caching context," *IEEE Access*, vol. 5, pp. 4018–4027, 2017.
[25] D. Chen, Z. Qin, X. Mao, P. Yang, Z. Qin, and R. Wang, "SmokeGrenade: An efficient key generation protocol with artificial interference," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 11, pp. 1731–1745, Nov. 2013.
[26] Q. Jiang, S. Zeadally, J. Ma, and D. He, "Lightweight three-factor authentication and key agreement protocol for Internet-integrated wireless sensor networks," *IEEE Access*, vol. 5, pp. 3376–3392, 2017.
[27] D. Chen *et al.*, "S2M: A lightweight acoustic fingerprints-based wireless device authentication protocol," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 88–100, Feb. 2017.
[28] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, 2014.
[29] C. Dwork, "The differential privacy frontier," in *Proc. Theory Cryptogr. Conf.*, 2009, pp. 496–502.
[30] J. Zhang, G. Cormode, C. M. Procopiuc, D. Srivastava, and X. Xiao, "PrivBayes: Private data release via Bayesian networks," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2014, pp. 1423–1434.
[31] N. Li, W. Qardaji, D. Su, and J. Cao, "PrivBasis: Frequent itemset mining with differential privacy," *Proc. VLDB Endowment*, vol. 5, no. 11, pp. 1340–1351, 2012.
[32] C. Dwork, "Differential privacy," in *Encyclopedia of Cryptography and Security*. New York, NY, USA: Springer, 2011, pp. 338–340.
[33] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. Conf. Theory Cryptogr.*, 2006, pp. 265–284.
[34] R. Bhaskar, S. Laxman, A. Smith, and A. Thakurta, "Discovering frequent patterns in sensitive data," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2010, pp. 503–512.
[35] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *Proc. 48th Annu. IEEE Symp. Found. Comput. Sci.*, Oct. 2007, pp. 94–103.
[36] C. K. Chui, B. Kao, and E. Hung, "Mining frequent itemsets from uncertain data," in *Proc. Adv. Knowl. Discovery Data Mining (PAKDD)*, 2007, pp. 47–58.
[37] M. Srivatsa, L. Liu, and A. Iyengar, "EventGuard: A system architecture for securing publish-subscribe networks," *ACM Trans. Comput. Syst.*, vol. 29, no. 4, 2011, Art. no. 10.
[38] Q. Wang, D. Chen, Z. Ding, Z. Qin, and Z. Qin, "MESA: An efficient matching scheme in content-based publish/subscribe system with simplified Bayesian approach," presented at 3rd Int. Conf. Big Data Comput. Commun., Chengdu, China, 2017.
[39] N. Zhang *et al.*, "Software defined networking enabled wireless network virtualization: Challenges and solutions," *IEEE Netw.*, to be published, doi: 10.1109/MNET.2017.1600248.
[40] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, 2008.

QIXU WANG was born in Neijiang, China, in 1985. He received the bachelor's degree from the School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang, China, in 2009. He is currently pursuing the Ph.D. degree with the School of Information and Software Engineering, University of Electronic Science and Technology of China. His research interests include information security, cloud computing and storage, and wireless network security.



DAJIANG CHEN (M'15) received the B.Sc. degree from Neijiang Normal University, Neijiang, China, in 2005, the M.Sc. degree from Sichuan University, Chengdu, China, in 2009, and the Ph.D. degree in information and communication engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, in 2014. He is currently a Post-Doctoral Fellow with the University of Waterloo, Waterloo, ON, Canada, and also with the School of Information and Software Engineering, UESTC. His current research interests include wireless security, physical layer security, information theory, and channel coding and their applications in wireless network security and wireless communications.



NING ZHANG (S'12–M'14) received the B.Sc. degree from Beijing Jiaotong University, Beijing, China, in 2007, the M.Sc. degree from the Beijing University of Posts and Telecommunications, Beijing, in 2010, and the Ph.D. degree from the University of Waterloo, Waterloo, ON, Canada, in 2015. He is currently an Assistant Professor with the Department of Computing Science, Texas A&M University-Corpus Christi, Corpus Christi, TX, USA. Before that, he was a Post-Doctoral Fellow with the University of Waterloo. His current research interests include dynamic spectrum access, 5G, physical layer security, and vehicular networks.

ZHE DING was born in Lanzhou, China, in 1982. He received the B.S. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, in 2007, and the M.S. degree from Lanzhou University in 2012. He is currently pursuing the Ph.D. degree with the School of Information and Software Engineering, UESTC. His research interests include machine learning and recommendation algorithm.

ZHIGUANG QIN (S'95–A'96–M'14) is the Director of the Key Laboratory of New Computer Application Technology and the Director of the IBM Technology Center with University of Electronic Science and Technology of China. His research interests include wireless sensor networks, mobile social networks, information security, applied cryptography, information management, intelligent traffic, electronic commerce, distribution and middleware, and so on. He served as the General Co-Chair for WASA 2011, Bigcom 2017, and so on.