IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 2, 2020, accepted November 14, 2020, date of publication November 17, 2020,
date of current version December 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3038825

A New Method With Swapping of Peers and Fogs to Protect User Privacy in IoT Applications

MOHAMMAD YAMIN[1] AND ADNAN AHMED ABI SEN[2]
[1]Department of MIS, Faculty of Economics and Administration, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[2]Islamic University of Madinah, Medina 42351, Saudi Arabia

Corresponding author: Mohammad Yamin (myamin@kau.edu.sa)

This work was supported by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, Saudi Arabia, under Grant FP-172-42.

ABSTRACT Internet of Things (IoT) is now serving as a vehicle to a huge number of applications, resulting in innovative and smart solutions in many fields. While providing abundant benefits which improve the quality of our lives, the IoT environment has also created new challenges, especially to users' privacy. Many IoT applications use location-based services (LBS), where service provider (SP) trust cannot be taken for granted. Sensitive information available to SPs could be used to cause considerable loss or damage to users' property or even endanger their lives. There are several methods to preserve the privacy of users' data from SPs but they all suffer from one or more anomalies. This research presents a new method, known as the Swapping of Peers and Fogs (SPF), to protect users' privacy from SPs. By exploiting the features of fogs and smart dummies, the SPF approach offers remarkable improvements to the level of protection of users' identity, which can be used to extract personal information. The SPF method does not compromise accuracy and, by using a pair of caches and fogs, provides greater efficiency for applications as compared to the existing approaches. To demonstrate the effectiveness and efficiency of the proposed method, detailed comparisons with current methods are presented via simulations based on different scenarios. Finally, an application of the SPF method to connected street systems in smart cities is also discussed.

INDEX TERMS Privacy, IoT, LBS, smart dummy, fog, peers, sensors, smart city.

I. INTRODUCTION

Internet of Things (IoT) now contains numerous innovative technologies in addition to billions of smart devices and objects [1] connected to the Internet. These smart things, spread everywhere around us, are helping to make our routine actions swift, flexible and sophisticated [2].

Many applications of IoT use location based services (LBS). Smart City is one of the natural and most powerful applications in IoT [3], which is heavily dependent on technologies like wireless sensor networks (WSNs) and RFIDs [4]. The job of WSNs is to sense environmental conditions, which can involve numerous variables like pressure, heat, noise, pollution, humidity, lighting, movement, leakage, sounds, images and so on [5]. When combined, WSN and RFID transfer artifacts of a designated space into smart objects, which can be used to share data with other objects as well as human beings.

The associate editor coordinating the review of this manuscript and approving it for publication was Danping He.

As expected, these smart tools are also limited in terms of energy, storage, and processing capacity [6]. Smart City applications rely on giant data centers, service providers and Clouds to provide a suitable environment to store and process the data generated by smart devices and objects. Thus, most smart city applications (whether related to environment, society, energy, health, economy, transport, etc.) rely on cloud computing [7], [8]. Cloud computing processes store and analyze data about users, then try to discover new knowledge and features that help smart cities to improve these services, thus providing smarter apps which are better adapted to each user [9].

A. PROTECTING AND PRESERVING PRIVACY

Preservation of security and privacy in smart cities has now assumed greater prominence as it poses an ominous threat to the future of smart applications and objects. As these smart things often work with and transmit sensitive data to faraway locations (clouds), sensitive data is susceptible to being hacked. In some situations, sensitive information in the

TABLE 1. Privacy vs. Security.

Privacy Prevents	Security Protects
Traceability	Confidentiality
Linkability	Integrity
Identifiability	Availability

wrong hands can cause devastation to users or the people associated with the information [10], [11].

A service provider (SP), normally a cloud, is always in an advantageous position to garner a significant amount of information about habits, behavior, personality, mindset, and ambitions of users or customers [12]. For example, while searching for points of interest, the location based services such as Smart Street, Smart Car, Ubiquitous Health, and Smart Alert enable the SP to track the user's location and find their whereabouts at a given time [13], [14]. An attacker can use the location of the user to gain vital information such as a user's identity, the whereabouts of their home location, and habits, in addition to the details of their job, religion, social leanings, and health data. As the number of IoT devices continue to climb, so does the vulnerability of data intrusion [15]. Thus, the protection of users' ID and location from the SP is crucial to preserve the privacy of users.

B. PRIVACY: THE FOCUS OF OUR RESEARCH

This article mainly addresses issues with the privacy of users from the SP in a LBS environment. Sometimes the concepts of privacy and security can get mixed-up, although they are different. As shown in Table1, privacy seeks to prevent miscreants to trace, link or identify users' personal data. On the other hand, security protects confidentiality and integrity of data and availability of services for applications. In view of [16]–[18], privacy is ''The right of the user to determine when, where, why, how, and who can access and use their data''. Protection of privacy in many situations is a complex problem but could be achieved by using methods to hide the identity from attackers or malicious parties such as the SP in the LBS environment, to prevent users from being profiled [19]. In other words, when sending a query to the SP, the user must ensure that the SP is denied access to any link which could provide them any useful information about them [20].

C. CONTRIBUTIONS AND STRUCTURE OF THIS PAPER

In this article we present a new method to protect the privacy of users by utilising a multi-swapping scheme in the LBS space involving peers and fog nodes (fogs). We call this method the 'Swapping of Peers and Fogs' (SPF). Details of the SPF method, including the swapping scheme and its justification, advantages, resilience, novelty and limitations are provided in Sections 3, 4 and 5.

In section 2, a literature review is presented, wherein Cloud and Fog computing are discussed. These technologies play a pivotal role in the SPF method. Also, a brief discussion of the strengths and weaknesses of the existing methods is provided, which are vital for comparing these methods with the SPF.

In section 3, we formally introduce the SPF method and offer justifications for its swapping scheme. The SPF method can be used in a variety of situations depending on the user requirement, which we analyse in section 4, In section 5, we present two algorithms to summarise the operational details of the SPF method. The first of these demonstrates how the swapping scheme works, and the second describes the cache management.

In section 6, we compare swapping schemes of SPF and other methods, discuss superiority, resilience, and limitations of the SPF method in detail. In section 7, we provide an example of deployment of SPF in the case of connected vehicles, as well as an application of SPF in a Smart City. In section 8, we analyse management issues associated with the swapping scheme, including the estimation of delay, disruptions, and the routing scheme of the SPF method. In section 9, by means of a set of performance metrics, we analyse swapping scheme of SPF and compare it with the other methods. In section 10, we provide simulations to compare the performance of the SPF method with the existing methods, including the details of the experiments.

Acronyms used throughout this article are presented in Table2.

TABLE 2. List of acronyms/abbreviations.

Acronym/abbreviation	Full Form
BTP	Blind Third Party
C_i	ith cell or cluster
C#	C Sharp
DB	Database
DOA	Double Obfuscation Approach
E	Entropy
Fog	Fog node
H	Cache Hit Ratio
IoT	Internet of Things
LBS	location Based Services
N_i	ith Fog node
P2P	Peer to Peer
P2PCache	Peer to peer cache
Peer	User in LBS
PIR	Private Information Retrieval
PID	Peer ID
P_i	ith Peer
Q	Query
RFID	Radio Frequency Identification
SP	Service Provider
SPF	Swapping of Peers and Fogs
TP	Third Party
TTP	Trusted Third Party
WSN	Wireless Sensor Network
U	Ubiquity

II. LITERATURE REVIEW

In this section we present a list of technologies and tools, and offer justification for using them. We also provide summary of some existing privacy methods, including their weaknesses

TABLE 3. Characteristics of Clouds.

Factor	Fog	Cloud
Nodes Number	Large number of nodes, which can cooperate in different forms (Cluster, P2P, and Master-Slave)	One or few servers
Storage Type	Caching data for a few hours	Permanent storage for data
Cooperation	Mostly cooperative (hierarchically) but the fog work independently	Mostly independently, but it can can cooperate with another cloud
Connection	Wireless connection	Internet connection
Location	At the edge of the network close to the user	Far From User
Distribution	Densely distributed	Central
Application Type	Supports new types of applications requiring high speed, and RT interactivity	Applications require powerful computing and big data analytics
Real Time	Strongly Supported	Weakly Supported
Acting as a broker	Can play this role between users and SP	It is the main party
Mobility	Strongly Supported	Weakly Supported
Accountability	Weak, but that increases the privacy of users by hiding their IDs	Strong, but that creates more threat to privacy

and relative performance. These methods are are referenced throughout several times in this article.

A. CLOUD AND FOG COMPUTING

Cloud and fog play critical role in the SPF method. The service provider, SP, is a cloud, and fogs are used in the swapping scheme. The number of devices that are connected to the internet is estimated in the billions [21]. With so many devices, cloud computing can no longer provide prompt response to the huge amount of smart applications, especially medical Apps, which are sensitively dependent on time [22], [23]. To meet such requirements, many solutions such as mobile clouds or multiple clouds and fog computing emerged [24] in 2012. Fog computing, with far superior features than cloud computing, is part of the solution to provide a faster response.

In typical applications, fog nodes (which we shall simply call 'fogs') are widely distributed at the end of the network and IoT devices (perception layer), which is closer to the user [25]. This setup manages a cluster or region with tools to provide responsiveness, especially in emergencies, as well as initial processing of data before sending it to a cloud for permanent storage. Fogs can store data for a short period of up to two hours, which is usually enough for nodes to collect and summarize the data [26]. The next step for the fogs is to send the data directly to the cloud, eliminating the need for hundreds of connections from numerous devices to interact with the cloud every few seconds. A batch transmitter also has a significant role in reducing load and improving the performance and privacy dramatically within applications that use this technology [27]. Fog nods (fogs) are made up of a hierarchical structure and share information with the core fog. The core fog is headed by the cloud, which is a distributed structure instead of a centralized one [28]. The main differences between fog and cloud structure are summarized in Table3. For more details, refer to articles [25]–[27], [29]. From Table3, it is evident that fog computing cannot be a substitute for cloud computing, but with their integration, a higher level of services, applications and features can be provided. A user query, when submitted to the LBS, has a number of components, which are shown in Table4. Rapid and massive growth in the number of objects of IoT, spread all around us and the Cyber Space, has resulted in heightened security and privacy concerns [30], [31]. Many of the existing protection techniques rely on the SP as a trusted party, and only focus on external attackers. Since the SP cannot be trusted, any privacy technique reliant on the trust of the SP is not reliable. However the trust of the SP is not critical in the application of data security protection methods, if transmission occurs and nicknames for users are used. Accordingly, advanced methods based on the trust of the SP should have an inbuilt system to alert users to grant permission for their data to be accessed [32]. How to avoid dependence on the trust of the SP has been an open problem, which has recently been addressed in the Blind Approach [33] by way of using a pair of keys in addition to the third party, and in the Double Obfuscation Approach (DOA) [34].

TABLE 4. Constituents of submitted queries.

Attributes of a Submitted Query				
Object Type	Data/Query	Identity(ID)	Location (Loc)	Time

B. EXISTING APPROACHES, THEIR STRENGTHS AND WEAKNESSES

Review existing approaches with a specific aim of highlighting their strengths and weaknesses.

There are many approaches and methods to preserve privacy but most of them suffer from one or more anomalies [16], [35]. Moreover, some of the existing approaches have given rise to challenging issues and open problems concerning performance, trust, and the impact of the core service and applications they provide. Most of the existing

TABLE 5. Performance of privacy methods (Criteria: ✓-Fulfilled; – - In balance; x - Not Fulfilled).

	Criteria													
Privacy Methods	No overhead on user	No overhead on server	No overhead on the link	Method does not trust SP	Method does not trust TP	Method does not trust Peers	Protect Location	Protect Data or Query	Protect Identity	Accuracy of the results	Protect privacy on server	Protect privacy on transmission	Protect privacy on user device	Protect data from Server itself
Data Encryption	–	–	✓	x	✓	✓	x	✓	✓	✓	✓	✓	–	x
Blind Approach	–	x	✓	✓	✓	✓	–	–	✓	✓	✓	✓	–	✓
TTP	–	–	x	✓	x	x	x	x	✓	✓	✓	x	x	–
Dummies	–	x	x	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓
Obfuscation	–	–	x	✓	✓	✓	✓	x	x	x	✓	✓	x	–
PIR	x	x	x	✓	✓	✓	✓	✓	x	–	✓	✓	x	✓
Cooperation	–	✓	–	✓	✓	x	–	–	–	✓	✓	✓	x	✓
Caching	x	x	✓	✓	x	✓	–	–	–	✓	–	–	–	–
DOA	–	–	x	✓	✓	✓	✓	–	✓	✓	✓	✓	–	✓
SPF	✓	✓	✓	✓	✓	–	✓	✓	✓	✓	✓	✓	–	✓

approaches depend on the trust of service providers, which is a major weakness and a serious deficiency. Performance of the main approaches against several criteria is summarised in Table5. Their description and weaknesses, which will follow, are relevant to our research in this article.

1) DUMMY APPROACH [36]

Purpose: The main purpose of using a dummy is to conceal the real query by mixing it with a set of dummy (unreal) queries to mislead the SP. This method can be used to protect the query or location. The SP will not be able to identify the actual query, and hence would be misled to collect inaccurate information about the users.

Hypothesis:

- Users are able to create dummy queries by themselves
- User resources enable them to create 30 dummy queries for every real query

Weakness:

- This approach causes overhead on the user as well as the SP as the number of dummy queries grows.
- After observing for a while, the SP can distinguish the user from others.

2) OBFUSCATION APPROACH [37]

Purpose: In this approach, the combination of the query and data of the user is changed before it is sent to the SP, unlike having to send many queries in the dummy approach. The level of privacy is related to the amount of noise and obfuscation on the query. Privacy can be increased at the cost of accuracy of results.

Hypothesis:

- Users are prepared to sacrifice the accuracy of results to protecting their privacy
- User has enough resources to recover the returned result.

Weakness:

- Increasing privacy would also increase the cost of processing. Newer Obfuscation techniques require the user to send their area instead of the location. But this method also adversely affects performance and cost. More importantly, Obfuscation is not suitable for smart street applications as it changes the locations of vehicles.

3) DOUBLE OBFUSCATION APPROACH [34]

Purpose: Double Obfuscation Approach (DOA) is a recent hybrid method to protect the privacy of users in LBS applications. It depends on obfuscation and Fog as the third party (TP) to enhance privacy compared to the traditional obfuscation, and addresses some drawbacks related to overhead and accuracy of results in the Obfuscation Approach [37]. To achieve that, it bifurcates the obfuscation area (one for the user and another for fog), and divides the returned results into five parts with the help of fog.

Hypothesis:

- same as in the case of the Obfuscation Approach, with additional overhead for processing.

Weakness:

- The DOA applications results in overheads on the user and server, and the approach does not provide adequate protection for the data of the query.

4) PRIVATE INFORMATION RETRIEVAL [38], [39]

Purpose: Private Information Retrieval (PIR) provides privacy by utilising a large amount of data from the SP.
Hypothesis:

- This method assumes that the user can access a huge amount of data from the SP without the SP.
- Assumes that the user has resources to store information of the whole city and deal with it.

Weakness:

- Accessing a huge amount of data from the SP may not be feasible at all times.
- This approach is not practical to use with smart devices of IoT, which are scarce resources.
- Some PIR techniques use encryption.

5) COOPERATION AMONG PEERS [40]

Purpose: The main goal of this approach is to reduce the number of contacts with the SP. In this approach, each peer in the same cell seeks the answer of their query from other peers, before sending it to the SP. In other variations of this method, peers collaborate with each other and send the same query to the SP to prevent profiling.
Hypothesis:

- Assumes that there are many users in each cell and most of them agree to send the same data to the SP.

Weakness:

- This is not suitable with smart street services.

6) CACHE APPROACH [41]

Purpose: This approach is similar to other approaches in caching some queries' answers, and reusing them to respond to future queries.
Hypothesis:

- Assumes that there is open access point with self-management for storing the result of previous queries of users.

Weakness:

- This method is effective only when the cache-hit ratio is increased, which is proportional to the privacy and performance of the query.

7) BLIND THIRD PARTY PEERS [33]

Purpose: The BTP encryption is used by the Blind Approach, and its role is to change the identities of users.
Hypothesis:

- In this approach, the user avails all the benefits of using a third party (peer) without having to reveal any data to them.

Weakness:

- There is a possibility of collusion between the third party (BTP) and the SP to breach users' privacy.
- Encryption may cause overload on some users devices.
- BTP encryption usually results in more power consumption by users' devices.

8) GENERAL DATA PROTECTION REGULATION

General Data Protection Regulation (GDPR) is an initiative of the European Union (EU), which came into effect on 25th May 2018. The purpose of enacting GDPR is to empower users and clients' privacy by making it mandatory for service providers, who collect or manipulate data, to only access privacy information with prior permission from users or in a number of defined emergent situations. Moreover, the access control of the GDPR enables users to access and manage their data in the SP database, providing a simple solution to protect their privacy. However, after its promulgation, the GDPR has been facing compliance issues. Many service providers are not enforcing it the way it should be enforced. Moreover, so far, no one has been penalised for not enforcing or misusing the GDPR. As a result, many malicious parties are still able to violate user privacy in their domain. In other words, despite being a very good initiative, the GDPR is not effective enough, and hence we still need other means and methods to ensure the privacy of users. More details, including the principles of the GDPR, are available in [42]–[44].

III. A NEW METHOD TO PROTECT PRIVACY IN IOT APPLICATIONS

Use of Location-based services (LBS) is widespread. Amongst others, these services are used in smart cities and IoT applications including smart streets, self-lighting, pedestrian safety, smart parking, connected cars, medical applications and emergency response, congestion handling, alerts and warnings for drivers and other road users, remote surveillance, location related advertisements, search for points of interest, and smart signals.

In this article we propose, and provide details of, a new method to preserve users' privacy in IoT applications. We call this method the "Swapping of Peers and Fogs" (SPF). The LBS environment can be regarded as a set of clusters/cells, each having a fog and a number of users (peers), who can cooperate with each other. With the SPF, Swapping can take many forms involving peers and fogs. Ideally, a user, P_1, would send their query to a peer, P_2 (surrogate peer), who would relay it to fog node (fog), N_1, in cluster C_1, who would send it to another fog, N_2, in cluster C_2. Then N_2 would send it to another peer, P_3 (submitting peer), in C_2, who would submit it to the LBS. Indeed, the main purpose of the swapping scheme of the SPF method is to temporarily change the Ownership (Identity) of the query using a surrogate and submitting peers, successively. The result of the query is sent back to P_1 by the same route in the opposite direction. In this scenario, it is evident that the privacy of P_1 would be protected from the SP, N_1, N_2 and P_3 but not from P_2.

A. JUSTIFICATION FOR THE SWAPPING SCHEME

A number of existing privacy methods use some form of swapping involving peers. The P2PCache Approach [45] uses a swapping scheme between the user and another peer in the same cell to create smart dummy (a query submitted by a peer

on behalf of the owner), to change the identity of the owner of the query before sending it to the SP. This is very similar to the swapping between the user and surrogate peer in the SPF method, except that in the P2PCache the same peer acts as a surrogate as well as submitting peer. As a result, the user and the cooperator peer would be located in the same cell, making the user's location vulnerable. Incidentally, none of the methods, which were evolved before the SPF method, resolved this issue.

Indeed, the swapping scheme of the SPF addresses the issue of submitted location. When a query is submitted to the LBS, the SP will not get information about the real owner, P_1, and instead would receive misleading information (query and location) about the submitting peer P_3. So, there is no way for the SP to find out any information about P_1, who would be quite far from P_3 and would appear to be in a different cell to the SP. Existing approaches (Dummy [46], Obfuscation [47], [48], PIR [39], and Cooperation [49]) have also used some techniques to preserve privacy but each of them has created serious issues. These issues are successfully resolved by the swapping scheme of the SPF method.

In section 8, we analyse a number of aspects of the swapping scheme used in the SPF method, including efficiency, extent of the processing delays, management of the processing in case a participating peers leaves the LBS area without completing their assignment, and the estimates of times clarification on the forward and backward routing.

IV. DIFFERENT SCENARIOS OF SWAPPING IN THE SPF METHOD

The SPF facilitates several combinations of swapping between peers and fogs. In an ideal situation, the first swapping would occur between the user and the surrogate peer to convert the real query of the user into a smart dummy in the same cell. In the second swapping, the surrogate peer (smart dummy) would transfer the query to a fog in the same cell. The third swapping would occur when one fog transfers the query to another fog, which in the fourth swapping would transfer it to the submitting peer. This is only one scenario of swapping. There are several other possibilities of which we only describe five. These scenarios take into consideration the possibilities of extraordinary situations at the time of applying the method. On the other hand, they demonstrate that the SPF method is flexible and adaptive. At any given time, a user might face one of the following (rare, but possible) situations, some of which will be taken into consideration in the discussion of the five scenarios.

- In an unlikely situation, a user is alone in the cell, who has to connect to the SP directly.
- A peer in the same cell in some situations (flat-battery) is unable to cooperate.
- If a user does not trust peers in the same cell or trusts Fog more than peers, the query may be sent directly to the Fog.
- A user does not trust the Fog. In such a case, they can increase the cooperation amongst peers in the same cell.
- A user wants to enhance the level of privacy. In such a case, they may increase cooperation amongst peers at the expense of creating overload.

a: ***TASK:***

A user wants a query (Q) to be processed by the SP, but at the same time does not want to disclose the ID. location or query to the SP, the submitting peer or the tw fogs.

b: ***NOTATION:***

As the LBS area is divided into different clusters/cells, which we denote as $C_1, C_2, C_3, C_4, \ldots C_n$. Each C_i is provided with a Fog Node N_i, and may contain a number of Peers (users) $P_1, P_2, P_3, P_4 \ldots., P_n$ at a given time. Following are the different phases of swapping which can take place involving Peers and Fogs.

1) **First Swapping:** P_1 sends Q to another peer P_2 in order to hide the real ID of P_1 from N_1.
2) **Second Swapping:** P_2 sends Q to N_1 in a cell C_1.
3) **Third Swapping:** N_1 sends Q to N_2 in a nearby cell C_2.
4) **Fourth Swapping:** N_2 sends Q to a peer P_3 in C_2 (submitting peer), who submits it to the SP.

A. FIRST (MAIN) SCENARIO OF THE SPF APPLICATIONS

This is the best case scenario. As shown in Figure 1, it involves all four phases of swapping. This scheme of swapping will result in providing complete protection of privacy (ID, location, and query) to P_1 from the SP, the submitting peer and the fogs but not P_1, which we discuss in the next section. It should be noted that the location of the query Q will remain the same in all phases of swapping.

B. SECOND SCENARIO OF THE SPF APPLICATIONS

Figure 2 shows the second scenario, in which N_2 in C_2 does not have a suitable peer, and so N_2 sends the query to the SP directly. This would enhance performance but the privacy protection, compared to the first scenario, would be slightly less because the SP could be curious about the source of the query.

C. THIRD SCENARIO OF THE SPF APPLICATIONS

If there is no suitable peer in C_1 then P_1 can send Q directly to N_1, who would forward it to N_2, who would then assign it to P_3 to deal with the SP. In this scenario the privacy of P_1 would still be protected from the SP but N_1 can access some information about P_1 in its cell. Details are shown in Figure 3.

D. FOURTH SCENARIO OF THE SPF APPLICATIONS

It is possible that, during the processing of a query, a cooperator peer abandons the process and exits the LBS area without completing their task. If this happens, the user would have to start the process again, which would cost additional time.

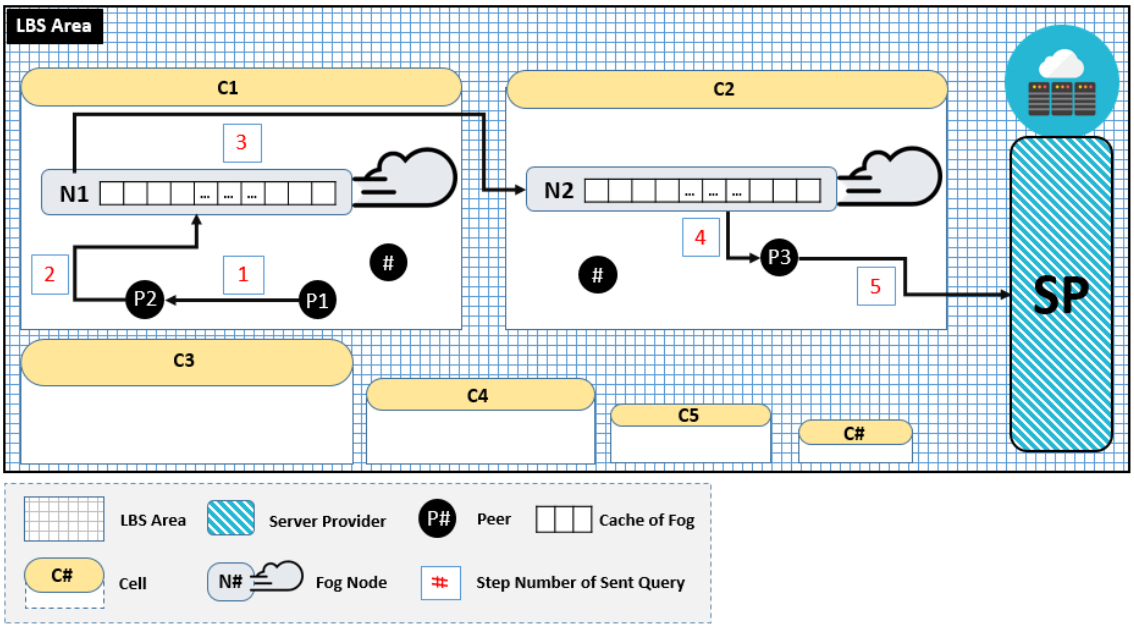

FIGURE 1. First (Main) Scenario of the SPF.

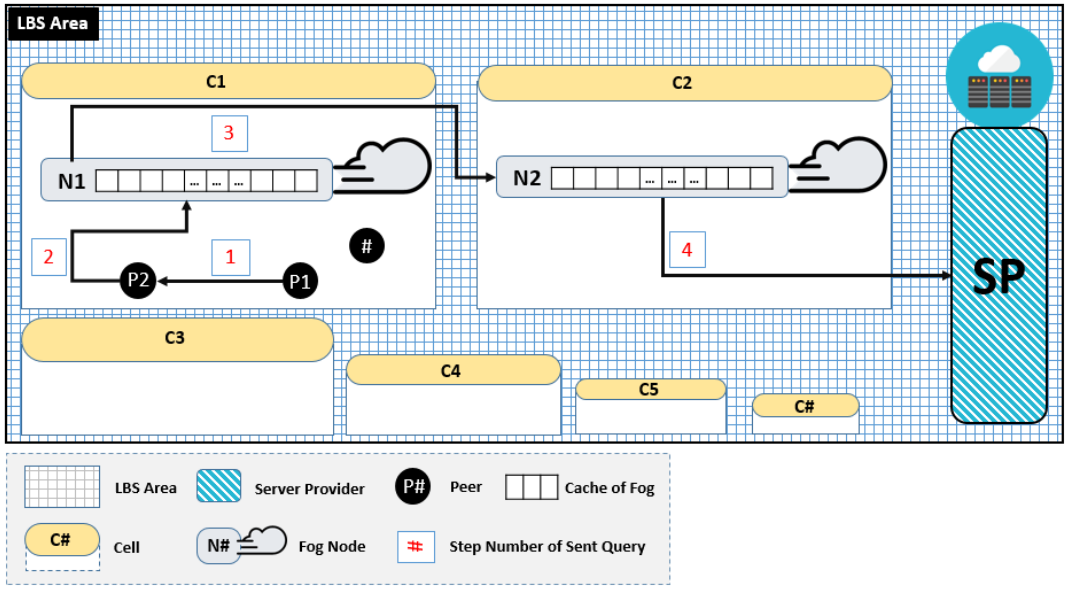

FIGURE 2. Second Scenario of SPF.

To avoid such a situation and ensure timely resolution, the user can send the query to two (surrogate) peers. In this case, there would be two parallel processes for the same query, as shown in Figure 4. It is highly unlikely that both of the processes would fail. Indeed, the duplication of the process would strain resources.

E. FIFTH SCENARIO OF THE SPF APPLICATIONS

In this case, P_1 trusts N_1 more than peers. Therefore, N_1 would manage the swap between peers. As shown in Figure 5, in this case only one fog N_1 would be used. As a result, location of P_1 would not be protected.

It should be noted that the submitting peer in a different cell (like P_3 in the First Scenario) boosts the level of their privacy (by misleading the SP about their location) without impacting the distribution of users in the cells/clusters of a smart city. So, the accuracy of the main services of a smart city such as Smart Street will not be affected after applying the SPF. This is one of the major differences between the SPF and earlier approaches to protect privacy. To avoid any adverse effect on

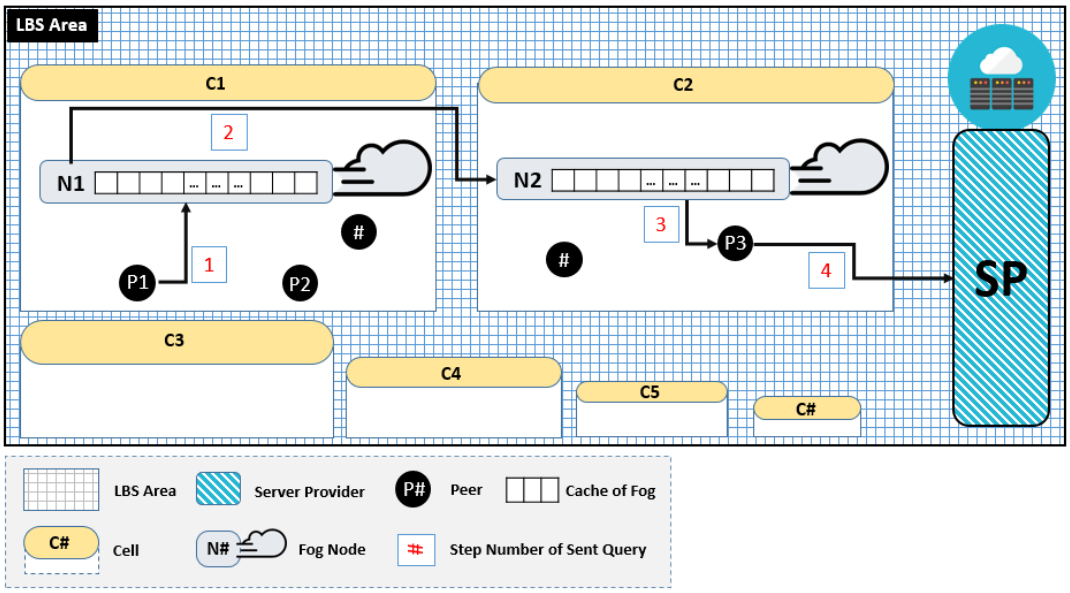

FIGURE 3. Third Scenario of SPF.

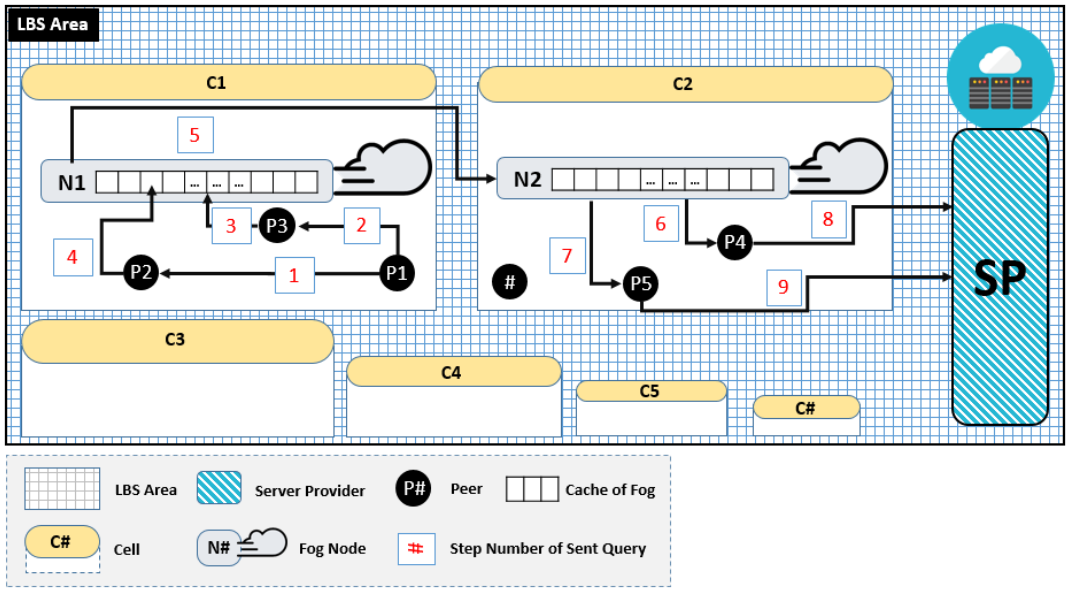

FIGURE 4. Fourth Scenario of SPF.

performance, the SPF uses the cache of fog node. The query usually passes through two fogs before contacting the SP. In other words, the effectiveness and higher cache-hit ratio are twofold advantages as compared to existing approaches which only deal with one cache. Use of fog in addition to clouds boosts efficiency for managing peers, caches, and other operations. Moreover, the SPF adds the bloom filter (Hash function) before searching in cache to avoid the miss-hit time of cache [50]. Furthermore, by default, the SPF only stores real queries (without dummies and noise in the cache) which enhances the cache-hit ratio.

V. ALGORITHMS FOR THE SPF METHOD

Here we provide two algorithms, which describe the processing of the SPF method. The first algorithm demonstrated the navigation of the swapping between peers and nodes, as described in the First Scenario (Figure 1), and is also included in the General Case of Vehicles (Figure 6).

The second algorithm describes the management of the pair caches inside the SP. As each fog has a cache, this algorithm makes use of them. We use a bloom filter (hash-table) to check if the query exists in the cache or not. If the answer is affirmative, we change the position and make this

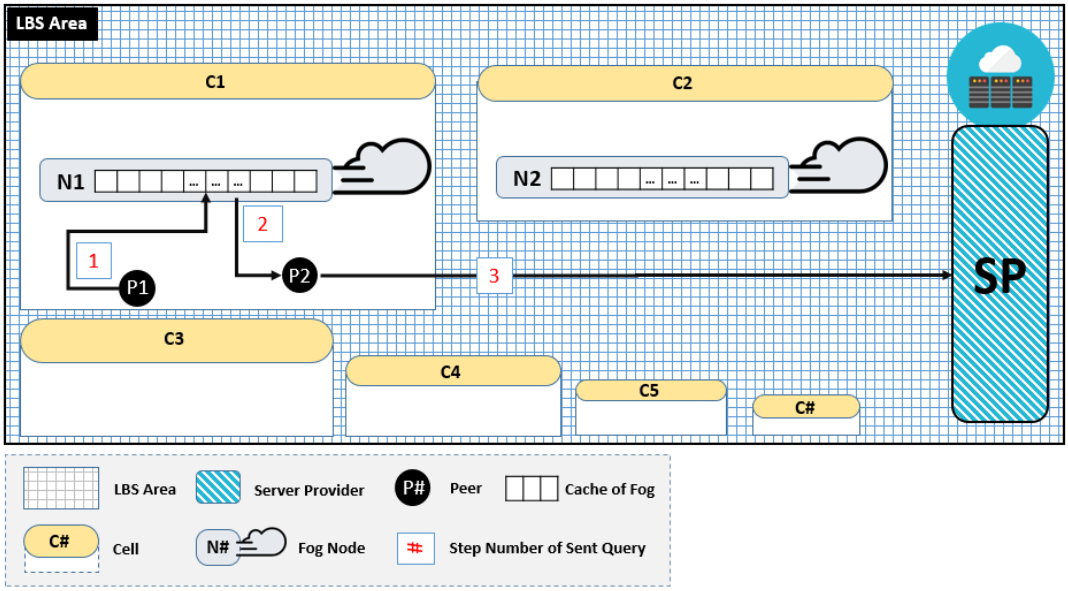

FIGURE 5. Fifth Scenario of SPF.

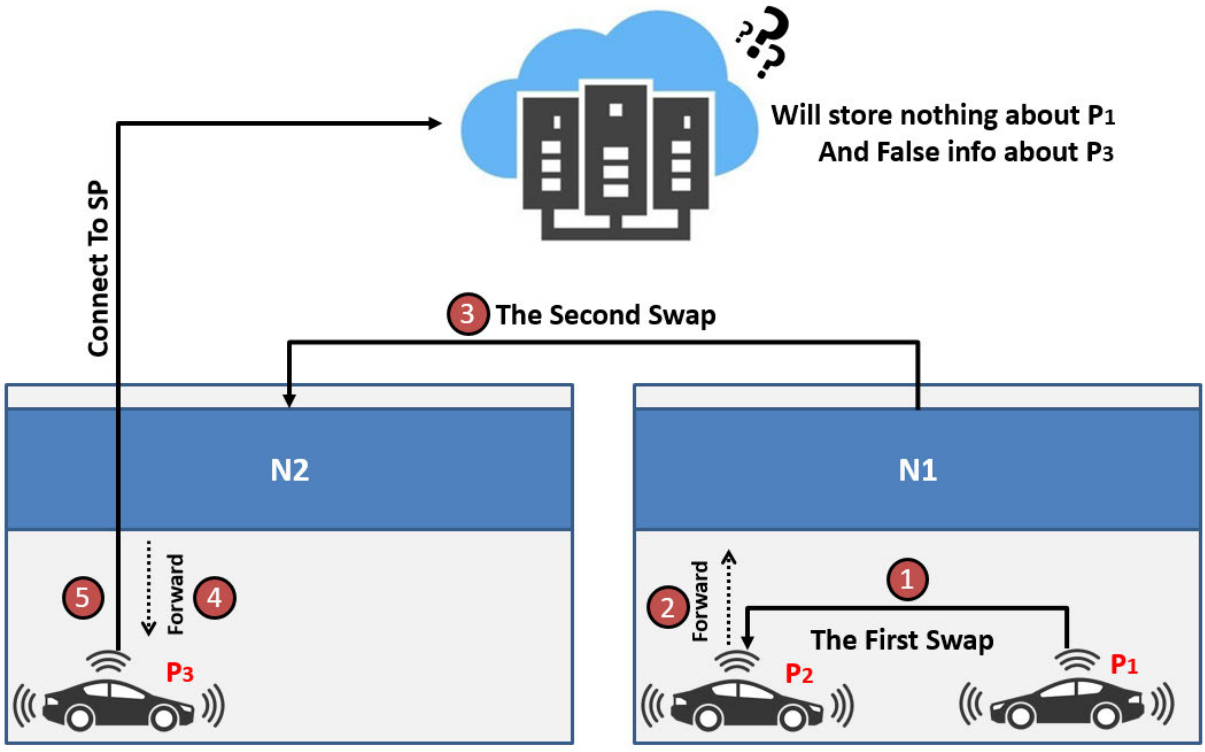

FIGURE 6. Simulation for main scenario of the SPF with Connected Vehicles.

as the first item in the list with MAX_{ID}. In case of a negative answer (miss-hit), we wait for the response of the query from the SP, upon the receipt of which we would delete the last query in the list (which has Min_{ID}), and insert the new one with $(MAX_{ID} + 1)$ in the first position. This process saves the items with a higher number of requests in the cache list.

VI. COMPARISON, ADVANTAGES, RESILIENCE AND LIMITATIONS OF THE SPF METHOD

The SPF method provides an efficient way to protect the privacy of the user's identity, location and the query from the service provider in the LBS environment such as smart cities. The level of protection is proportional to the number of users

Algorithm 1 First Algorithm: Swapping in the SPF Method

Input: Query Q // Query of P
Output: Results R[ ] // Resolution of Q
Initialisation:
P_1, P_2, P_3 = Null; N_1, N_2 = Null;
List LS, LP = New Peers [ ], LF = New Fog Node [ ]
Step 1
LF = GetListOfActiveFog(); // $(N_1, N_2 \ldots.)$
LP = GetListOfPeerFromMyFog(); // $(P_1, P_2 \ldots.)$
Step 2
if (N_1.LP > 1) **then**
  $P_2 = P_1$.Get-Random-Peer (N_1.LP);
  if LS has (P_2) **then**
    P_1.LP.Delete (P_2,); // this peer was used before
    Go to Step 2;
  else
    P_1.LS.Add (P_2,);
    R = P_1.Send (Q, P_2); // Send Q to P_2
    R = P_2.Send (Q, N_1);
    Step 3
    $N_2 = N_1$.Get-Random-Fog (LF);
    R = N_1.Send (Q, N_2);
    if (N_2.LP.Length>0) **then**
      $P_3 = N_2$.Get-Random-Peer (N_2.LP);
      R = N_2.Send (Q, P_3);
      R = P_3.Send (Q, SP);
      Break;
    else
      R = N_2.Send (Q, SP);
      Break;
    end if
  end if
else
  R = P_1.Send (Q, N_1);
  Go to Step 3;
end if
return R; // in the reverse order of sending.

Algorithm 2 Second Algorithm: For Cache Management

Input: Q = Query //Query (ID, Location, Type/POI, Range)
Output: R = Result of Q // R (Array of POIs and Locations)

Step 1 // Search-Function
if ((HASH-CACHE-Has (Q)) **then**
  Q_{ID} = (MAX_{ID} in $Cache_N$ + 1)
  Q_{ANS} = FETCH-Q-in-$Cache_{N()}$
  return Q_{ANS}
else
  R = P.Send (Q to SP);
  if (IsFull($Cache_N$)) **then**
    Delete Query with MIN_{ID} in $Cache_N$
    P.Insert Q_{ANS} to $Cache_N$ with MAX_{ID} + 1
  else
    P.Insert Q_{ANS} to $Cache_N$ with MAX_{ID+1}
  end if
end if
Go to Step 1

at the time of processing. The most important characteristic of this approach is the scheme of spreading fogs and using a pair of caches in smart cities, which are exploited to play a pivotal role in protecting the location of users from the service provider.

A. COMPARISON OF SWAPPING OF SPF AND OTHER METHODS

In case of the SPF, each submitting peer can only send one query, belonging to some other user. This mechanism eliminates the need to generate dummy queries or the user sending their own query to the SP. As a result, the SP would not receive any credible information about the user. On the other hand, the swapping schemes in the Obfuscation and P2PCache, who use Smart dummy, suffer from the following two anomalies.

1) Swapping between peers can happen only in the same area. In this way, approximate location of the user would be disclosed to the SP inadvertently by the cooperated peer.
2) There is no clear management of the relationship between peers in the area, an open problem in the cooperation approach.

The Swapping scheme of the SPF method removes these two anomalies in the following way.

1) The responsibility of managing peers is assigned in its cell to a fog.
2) The first (normal) swap between the user and surrogate peer in the same cell is used only to protect the privacy of the user from the Fog node of the cell (in case the node is malicious).
3) The second swap, which occurs between two nodes (of different cells), is the most critical in the SPF Method, because it would transfer the query from cell C_1 to cell C_2, who would send it to P_3. When P_3 submits the query to LBS, the SP would be disguised to record the ID of P_3 with the location and query of P_1 in C_1. In this way, privacy of the location of P_3 would also be protected, which is further discussed in section 6C.
4) The cache in each of the two fogs (double cache) would enhance the privacy and performance of the query.

B. SUPERIORITY OF THE SPF METHOD

The double swapping scheme of the SPF method encourages peers to cooperate with each other. In doing so, they boost their level of privacy. In particular, when P_3 submits the query of P_1, the query becomes a smart dummy, and hence the SP would record the wrong information about P_3. In other words, not only does the owner of the query, P_1, get full protection but the privacy of the submitting peer, P_3, also increases.

The first swap between P_1 and P_2 also removes the need to trust N_1. The SPF method also has the following advantages:

- Provides three tiers of protection for privacy, and there is no significant impact on performance.
- Addresses the problem of the dummy approach by creating smart dummies without creating an overload on the system.
- Solves the problem of the TTP approach by employing multi-fog nodes without having to trust them and the fogs facilitate a solution for the managing peers.
- Addresses the Double Obfuscation and Obfuscation Approach issues without affecting the accuracy of processed results or creating an overhead to the user.
- Provides a solution to the caching technique problem by improving the hit ratio (because the cache only contains real queries of users) and using the bloom filter with the help of an available pair of caches.
- Offers a solution to the cooperation approach problem by locating users in the same homogeneous area and within close vicinity.
- Provides resistance to most types of attacks.

C. RESILIENCE OF THE SPF METHOD AGAINST ATTACKS

Here we discuss privacy issues in different contexts, details of which are available in [16], [35], [47], [48], [51]–[53].

- **Semantic Context:** When an attacker has some additional information like the profession or the age of the user, then the attacker can use this information to break into the protection technique. However, in the SPF, the user would not deal with the SP at all, removing the possibility of such attacks.
- **Homogeneity Attack:** If the area of protection is homogeneous (has one stamp or same type of building), the obfuscation or cooperation among peers will not be useful. But the SPF uses additional swap among fogs to change the area of the user completely. Hence, a Homogeneity attack would not be successful in this case.
- **Path Tracking:** An attacker may try to draw a path for the users' positions by time (Historical data), to detect the direction and target of the user. This becomes easier with minor obfuscation, or cooperation among closed peers. However, the SPF uses swapping between two different areas with different nicknames, ensuring that the attack would also not be active in such cases.
- **Inversion Attack:** If an attacker has information about the protection method, they can access the privacy data by breaching the protection. However, in the SPF, despite having knowledge of using steps of the SPF, the SP can not link the received data to real users.
- **Knowledge of Map:** The attackers can use their skills or knowledge about the map to eliminate the dummy queries or noise from the obfuscated ones. However, in the SPF, real queries are known only to the real users, so the attacker can not do anything here.
- **Malicious Peer:** This is an open problem found in the cooperation approach. In the SPF, the nickname mechanism solves this issue, and dealings with the same peers rarely happens in the mobile objects environment.

In section 9, we provide formal analysis, which further elaborates the resilience of the SPF method.

D. LIMITATIONS OF THE SPF METHOD

There is no method which can effectively protect privacy from all nodes in the LBS environment, or serve all types of applications of IoT. Some of the available methods rely either on the trust of the SP or some other node, whereas others suffer from a range of operational anomalies. In case of the the LBS environment, the most dangerous node to breach the user privacy is the SP, followed by the two fogs (N_1, N_2). The least dangerous are the surrogate and submitting peers, (P_2 and P_3). In view of the preceding discussion, the SPF method protects user privacy from the SP, the two fogs, and the submitting peer but not from the surrogate peer. As a user randomly chooses a surrogate peer, it is highly improbable that a user would choose the same peer again to act as a surrogate. In general the chances of the surrogate peer being a heckler are slim. Nevertheless, the exposure of user's privacy to the surrogate peer is indeed a limitation of the SPF method. There are some other minor disadvantages of the SPF methods, which are listed below.

- There may arise, although very rarely, a situation when there are no peers in the application area. In such a case, the user should either randomly generate dummies or rely on the swap of the fog node only.
- If, for some reason, the surrogate peer P_2 leaves the LBS environment before the result of the query comes back, the user can choose another peer, which we discuss in section 8B in detail.
- Despite using fog and cache, the SPF process may cause delays in some user queries, although the effect would not be noticeable in the system overview. The delay can happen if there is no other peer in the same cell and the user decides to wait for a peer. Although rare, it would still be a possibility. Instead of waiting for a peer, the user can deal with the fog node directly like in the Third Scenario shown in Figure 3.
- The system focuses on the protection of privacy, not on ensuring the credibility of the processed results, which is an issue related to the reputation algorithms of both the service provider and the fogs responsible for providing services. If a service provider tampers with, it would immediately be detected and reported. As a result, the service provider would loose users after a short time. We shall deal with this issue in our future studies.

VII. APPLICATIONS OF THE SPF METHOD

In this section we present the general form of the case of vehicles, and then an application in Smart Streets.

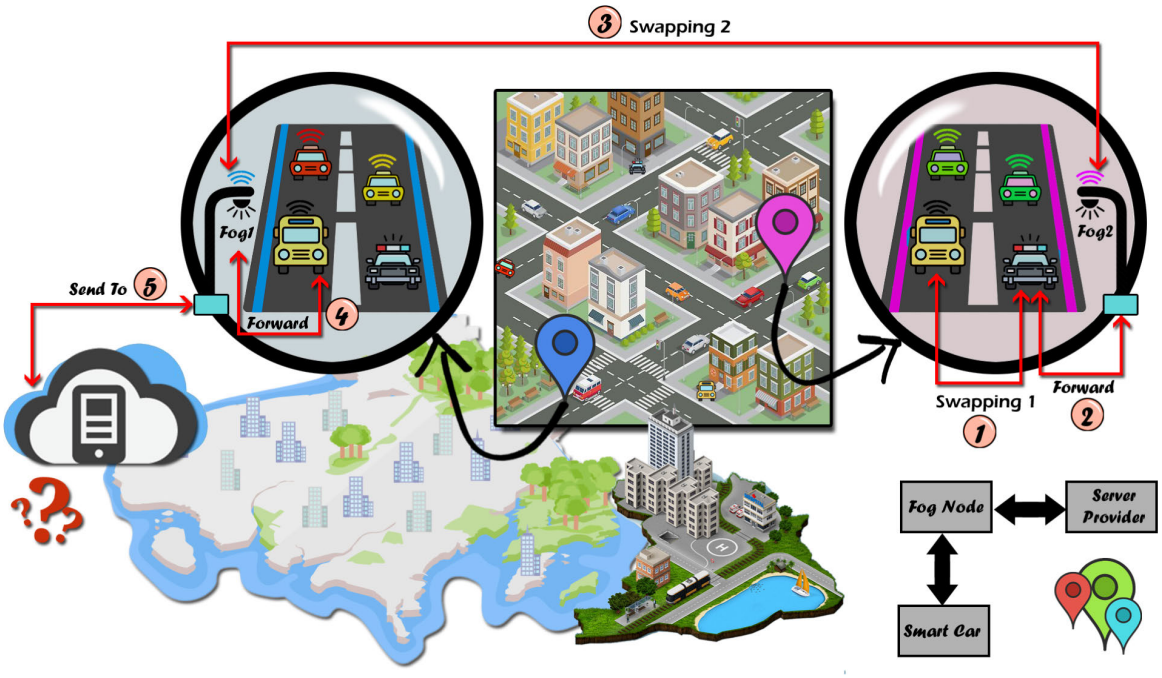

FIGURE 7. Example of deployment of the SPF in the Smart City.

A. GENERAL CASE OF VEHICLES

We first present the general form of the case of vehicles.

1) A vehicle with a query (Q) would generate a random nickname like P_1.
2) P_1 would swap its query with another vehicle P_2 in the same zone.
3) P_2 would send P_2.Q to a fog node N_1 in its cell.
4) N_1 would swap P_2.Q with another fog node N_2 in another cell C_2.
5) N_2 would assign P_2.Q to vehicle P_3.
6) P_3 would send P_3.Q to the SP.
7) The SP would return the results to P_3.
8) P_3 would forward it to N_2.
9) N_2 would provide the results to N_1 according to the ID of the query.
10) N_1 would return the results to P_2 which in return will return them to P_1.
11) Fogs would save the results in their caches to answer future queries without having to deal with the SP again.
12) With another query, A can deal with another peer and nickname it to prevent any possibility of linking data by time to vehicles.
13) Then the fogs would search in their cache before contacting the SP.

In the previous scenario, if an intruder (outer attacker or malicious SP, fog or Peer) traces any vehicle like P_1 by that time, then this attacker will create a false profile about P_1 and other vehicles, and the attacker will have random paths for each one (P_1, P_2, etc.). (See Figure 6)

B. AN APPLICATION OF THE SPF IN SMART STREETS

Smart streets are the most important applications provided by smart cities, where they include a large number of diverse services. The most important of these services include automated addressing of congestion and flow management, medical services everywhere, immediate response to emergencies, easy and accurate search for points of interest, self-lighting, safety of pedestrians, Smart parking, pollution sensors, noise and leaks, alerts and warnings for drivers, remote monitoring, advertisements associated with the place, search for points of interest, monitoring violations, and others [54].

A characteristic of all aforesaid services is that they depend on the location service. For example, in order to solve the problem of congestion in smart streets, the services might rely on automated and continuous calculations for the number of vehicles in each area, which is dependent on current locations of these vehicles, which connect with the SP who is responsible for guiding them to the less congested roads [55]. To protect privacy in this case, the first option is to rely on third-party trust, which is not a real solution. The second option is to rely on one of the dummies or jamming methods, which protect the location. But in this kind of application, the use of dummies would affect the number of vehicles, and the use of noise would affect the proportions of the distribution of vehicles in the region instead of the real proportions. It means that the protection technology has negatively affected the quality and effectiveness of the basic service related to congestion addressing.

The steps are shown in Figure 6, whose description follows: In the SPF, it will provide several benefits:

- Benefit from the presence of Fogs distributed in smart streets.
- Achieve protection of identity, location and query together.
- The protection approach does not affect the distribution rates of vehicles in each area due to the process of double swapping between the fogs where the same number will remain at each node.

VIII. MANAGEMENT OF QUERY PROCESSING AND PEERS IN THE SPF METHOD

In this section, we (a) provide an estimate of possible delay in query processing by the SPF scheme, (b) discuss the management of cooperated peers, and (c) explain the reverse routing process from the LBS to the user.

A. ESTIMATION OF POSSIBLE DELAY CAUSED BY THE SPF PROCESS

It is known that the searching in DB of the SP, namely (T_s) takes more time than in Cache (T_c), where in the worst case scenario, T_c = Time Access Cache × Number of Elements. However, for our calculations, we regard T_s and T_c to be the same, and so we do not take them into account in our comparison. Moreover, We use a bloom filter to avoid a search in the cache in case there is no result; which costs 1ms of time. However, if the query already exists in the cache, the search time is related to the size of cache (in our experiment, with a 100kB sized cache, the search time was 50ms).

The average time for a small size of information (like a Ping Test) to establish an online 4G connection with the SP is about 100ms (which may vary a little according to the speed of the connection), whereas a query, like Ping, for an offline wireless connection takes about 10ms. Using previous statistics, we can estimate the possible delay that would occur in the SPF process. In the main scenario of the SPF, we have four traversals in a wireless connection before the final internet connection is established with the SP. In this case, we have the following time estimates.

- Time taken by the query without protection will be T_1 = A (Example A = 100ms)
- Time taken by the SPF to process the query in a worst case scenario will be $T_2 = T_1 + 4 * B$ (Example B = 10ms)
- Time taken by the SPF process when the result was in the cache of Fog Node (1 or 2) would be $T_3 = 3 * B$
- If H = cache hit-ratio, the total time for processing N queries will be N*T_3*H + N*T_2*(1 − H)

To substantiate the forgoing discussion, we conducted a small experiment to quantify the delay caused by the SPF process in milliseconds (ms) by using a Ping Test (with a 4G network). We repeated the experiment ten times and calculated the averages of each count, which are provided as follows: Each of the first four traversals in the forward routing took 4B = 4*10 = 40ms (WiFi) and the last connection with SP took A = 40ms (Internet).Thus, the total time, denoted by TT, comes out as TT = 40 + 100 = 140ms. So, there is a 40ms delay after using SPF method for each query in case if the query results from the SP is communicated in a normal way. However, if the result exists in Cache, there is no need to connect to the SP, and hence the total time would be less than 100ms, resulting in no delay.

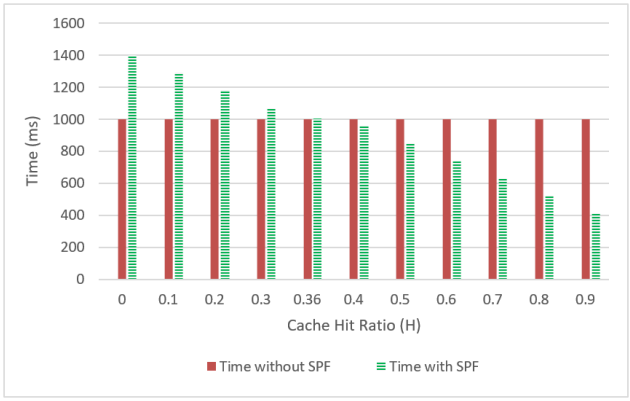

FIGURE 8. Time Estimates for Processing with and without SPF.

To address the issue of delay, the SPF method depends on the double cache of fog nodes. Figure 8 shows the total time for 10 queries with different H values. It should be noted that, in normal cases, if H = 0.36 then the SPF will not cause any delay, and if H < 0.36 then there will be some delay. On the other hand if H > 0.36, then the SPF will enhance the performance of the query.

B. WHAT IF PARTICIPATING PEER ABANDONS THE PROCESS MIDWAY?

In a real and dynamic environment, like the LBS and Connected Vehicles, there is a possibility that the surrogate or submitting peer leaves the environment before completing their part of the process. Such a situation creates a challenge to any privacy scheme. This problem is dealt with in the following two ways:

1) USE TWO SURROGATE PEERS

If the query is time sensitive, employ two surrogate peers, and hence duplicate the process in parallel. As described in the Fourth Scenario, and shown in Figure 4, it would ensure a timely response.

2) MANAGE WITH A PAIR OF CACHES

This is a way to manage the situation by exploiting a pair of caches, $Cache_1$ of N_1 and $Cache_2$ of N_2, in the following four cases:

1) If P_2 abandons before sending query to N_1
2) If P_2 abandons after sending query to N_1
3) If P_3 abandons before submitting the query for processing

4) If P_3 abandons after submitting the query for processing

Let T be the average time to send the query for processing to the SP, and receive a response back. The value of T would depend on the type of application and environment. To determine T in the above four cases, let $Tcache_1$ be the average time to send a query to N_1 and receive the response back.

1) **First case:** If P_2 leaves the LBS area before submitting the query to N_1, then P_1 has to send the query again to another peer. The new response time in this case will be $NT_1 = T + T$.
2) **Second case:** If P_2 leaves the LBS area after submitting the query to N_1, like in the first case, the user should resend the query to another peer. However, in this case the time would be $NT2 = T + Tcache_1$, which is less than NT_1, because the result would already be available in $cache_1$ of NT_1.
3) **Third case:** Peer P_3 would rarely abandon the query before sending it to the SP, because NT_2 would select P_3 by monitoring the entry time to C_2. But if this does occur, then NT_2 will send the query to another peer in C_2, save the the query resolution in $cache_2$, and return the response to $cache_1$ of N_1. So $NT_3 = NT_2 = T + Tcache_1$, just like in the second case.
4) **Fourth case:** It is similar to the third case, and N_2 would deal with it accordingly. So, $NT_4 = NT_2 = T + Tcache_1$.

To summarise the forgoing discussion, if a cooperator peer leaves the LBS area before completing the task, the additional delay needs to be taken into account, with the following two possibilities:

1) If the query is not sent to the SP, and therefore the user doesn't get the result back after 140ms, then they need to send the query again, in which case TT $= 140 + 140 = 280$, resulting in net delay of 180ms.
2) If the query is sent to the SP and the response arrives to NT_2, but the user doesn't receive the result back after 140s, then they need to resend the query. This time, the result would already be in $Cache_1$, so TT $= 140 + 10 + 10 = 160$, recording a delay of 60ms.

In general, in any protection method, there is a trade-off between protection level and the processing time and performance/cost. In the future, we shall propose a novel idea to create a reputation for each peer, which would enable the user to only deal with trusted peers.

C. HOW WILL THE ROUTING INFORMATION BE MANAGED WHEN THE QUERY CROSSES THE PEERS, FOGS AND CELLS?

Backward routing of the processed query is the reverse of the forward routing, and both are listed below.

1) FORWARD ROUTING

$$P_1 \rightarrow P_2 \rightarrow N_1 \rightarrow N_2 \rightarrow P_3 \rightarrow SP$$

2) BACKWARD ROUTING

$$SP \rightarrow P_3 \rightarrow N_2 \rightarrow N_1 \rightarrow P_2 \rightarrow P_1$$

The peers in each cell will be managed by its fog, and each peer will have a different internal IP to connect only with other peers in the same cell (WiFi) or the fog itself to refresh the list of available peers. At the same time, each peer will have a special internet connection (like 4G) to connect to the SP directly. Each cluster can be managed by its fog, and all fogs can be managed be the admin of the smart-city.

In order to manage time, we use simulation (Packet Tracer) in addition to a real test on a small network by using "Ping" to check the time of sending and receiving the query, implemented as a short code by Visual Studio.NET. Also, we implement (ASP.NET C#) to manage Fog-Functions (Manage Peers, Swapping Q), and Search for the result in the cache by using a Bloom filter (Hash-Table).

IX. ANALYSIS OF THE SWAPPING SCHEME

To analyse the efficiency and effectiveness of the swapping scheme of the SPF method, we need some measurable metrics. Well known methods, namely the Dummy enhanced-CaDSA [36], Obfuscation [37], and Blind Third Party Encryption [33] have used the following six metrics to analyse their schemes. We shall also use the same metrics to analyse the scheme of the SPF method.

1) **K-Anonymity**: A measure of the extent (percentage) to which the SP is deceived by the scheme.
2) **Entropy (E)**: A measure of identifying the real location out of the anonymity set.
3) **Estimation Error (EE)**: Percentage of errors made by the attacker during their quest to determine the real query or location of the user.
4) **Cache Hit Ratio (H)**: Percentage of residual query in the cache. It is used for curtailing the number of connections to the SP.
5) **Cost-Time**: Total sending and receiving response time of the query (T).
6) **Cost-Size**: Size of transferred data (S).

In our analysis, we mainly focus on the user (P_1), and the Submitting peer (P_3), to calculate six metrics in relation to the SP as a perceived attacker, and compare them with the aforementioned three methods. First of all, we provide the values of the above metrics in the case of the query processing without any privacy protection method.

a: CASE OF QUERY PROCESSING WITHOUT PROTECTION

When there is no privacy protection, the metrics are as follows:

1) K-Anonymity $= 1/1 = 1$, which meas that SP understands that the query belongs to submitted peer P_3
2) E $= -\sum_{i=1}^{K} P_i log_2 P_i$, where P_i, the probability of query belonging to the submitting peer, is 1 and so E $= 0$.

3) EE == E ∗ 100% = 0, meaning no errors are made by the attacker (i.e the attacker knows who does the query belongs to).
4) H = maximum, because only real queries are saved in the cache
5) Cost-Time = T (time taken in sending the query to, and receiving the response from the SP).
6) Cost-Size = S (number of bytes sent to the SP for each query) = S(Q), where S(Q) == Identity(int32) + Latitude(double) + longitude(double) + POI (String) + Range(int16) == 4 + 8 + 8 + 50 + 2 = 72 Bytes, which is less than 1 KB for each query

From this analysis, it is evident that to enhance the privacy, K-Anonymity, E, EE, and H be increased, but T and S are not increased. These metrics in the above three methods are examined below.

b: CASE OF QUERY PROCESSING WITH DUMMY APPROACH

This approach [36] is based on creating many (K) dummy queries which are sent along with the real query. The six metrics can be summarised as follows:

1) K-Anonymity = 1/(1 + K), so increasing K would enhance the privacy
2) E would be maximum, if all the dummy queries are similar to the real query. In such a case all queries would have the same probability, and $Max(E) = log_2(K + 1)E$. But this cannot be achieved by the Dummy approach because it is very difficult to generate dummies similar to the actual query. In other words E would be enhanced but not to the extent of Max(E).
3) $EE = E*100\%$, showing dependency with the E value.
4) H would be smaller because the dummy data is stored in the cache, which will have an adverse effect
5) $Cost - Time = T * (K + 1)$, an indicator of adverse impact.
6) $Cost - Size = S(Q) * (K + 1)$, another indicator of negative impact.

c: CASE OF QUERY PROCESSING WITH OBFUSCATION APPROACH

This approach [37] is used to change the location (location privacy) of the user before sending the query for processing.

1) K-Anonymity = 1/D, where D is the distance between real and obfuscated location. So, higher privacy would be as a result of greater distance, which is raised at the cost of the accuracy of the result.
2) $E = -\sum_{i=1}^{D} P_i log_2 P_i$, D is the number of possible locations or number of previous queries of user received by the SP.
3) EE, is related to E
4) H, no impact
5) Cost-Time = Time of response for the query (T) + Time of mapping response to real location, impacting the processing time adversely
6) Cost-Size is not impacted

d: CASE OF QUERY PROCESSING WITH THE BLIND THIRD PARTY (BTP) ENCRYPTION

The BTP encryption is part of Blind Approach [33], and its role is to change the identities of users.

1) K-Anonymity = maximum = 0 because user does not deal with SP in a normal case.
2) E = MAX(E)
3) EE, maximum for the SP
4) H = 0 because encrypted data cannot be stored in the cache, which is an adverse impact
5) Cost-Time = 2*T (because of the connection with the BTP and then connection the BTP to the SP)
6) Cost-Size ≥ S, few additional bits to the last block if it is not completed, and the size of the sent key

e: CASE OF SPF

1) K-Anonymity = Maximum = 0 for P_1 and P_3, as P_1 does not connect to the SP, and P_3 submit a query belonging to some one else.
2) E = MAX(E) for P_1 because of no contact with the SP, while E for P_3 would be enhanced because K becomes K + 1 after each new query.
3) EE is Maximum because of E being maximum
4) Cache-Hit Ratio = H, because two caches are employed here, and only real queries are saved in the cache.
5) Cost-Time = T + Time of swap between P_1, P_2, N_1, N_2, and P_3, adverse impact. However, H will compensate this impact as discussed earlier
6) Cost-Size = S, No change

From the above analysis, we conclude that SPF is significantly superior in protecting privacy with only an insignificant impact on cost.

X. COMPARISON EXPERIMENTS AND SIMULATION RESULTS

Here we provide a comparison of the SPF method with those which use a dummy (Enhance-Cache) [36], cooperation among peers P2PCache [45], encryption and TP (BTP) [33], and Double Obfuscation Approach DOA [34]. All of these approaches also use the cache technique. In order to facilitate comparison, we use the following hypotheses which were used by these methods.

- The smart area contains a 100*100 cluster/cell, and each cell has a Fog node
- Each fog node has a cache
- The size of Fog's cache is 100K, while the size of one query is less than 1KB
- There are 10000 peers/customers who are spread randomly in the cells
- There are 100 POIs
- There are Wi-Fi connections (3G/4G Network)

The Performance Metrics consist of (a) number of queries sent to the server in each request, (b) percentage of time needed to process the query after and before being sent,

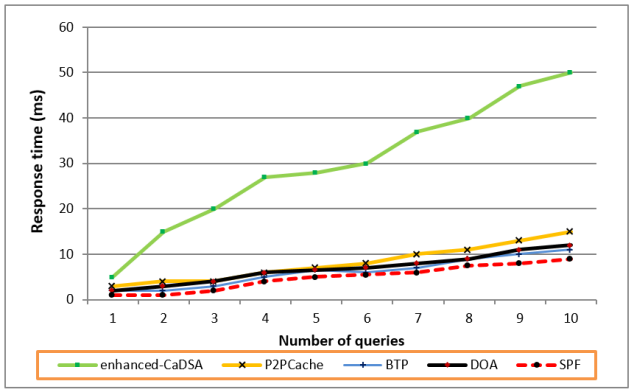

FIGURE 9. A Demonstration of Efficiency of the SPF.

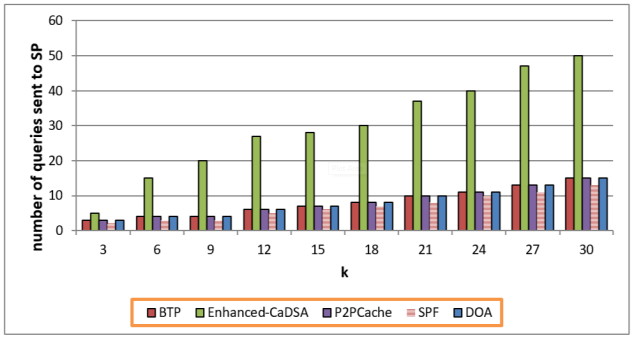

FIGURE 10. Comparison of the SPF with other methods.

and (c) Cache-hit Ratio (the number of queries that can be answered by the cache without needing to connect to the SP). Logic Metrics consist of (a) kinds of attacks that the protection technique is resilient to as discussed before, and (b) impact of the privacy approach to the core service. It should be noted that Privacy and Performance will be affected by the Cache-hit Ratio. Furthermore, the obfuscation or dummy can impact the number of vehicles in each area or street, which can adversely affect the applications of smart streets like the ones which are used to find the path of lowest congestion and/or traffic.

A. DETAILS OF EXPERIMENTS

Our simulation was carried out with the help of the Visual Studio 2015 (Asp.net C# and SQL Server 2012) and Microsoft Excel Office 365, in addition to the Cisco Packet Tracer Simulator. We wrote a code to conduct the experiments in accordance with the hypotheses. In order to analyse with performance metrics, we had generated random queries and took a part of data from the Geo-life dataset (which contains more than 17000 GPS paths for 182 clients over for three years). Then we applied our SPF Algorithm to check the time and the number of queries which were sent to the SP. To find Response Time, we repeated the experiment ten times for each query on different devices and then took the average.

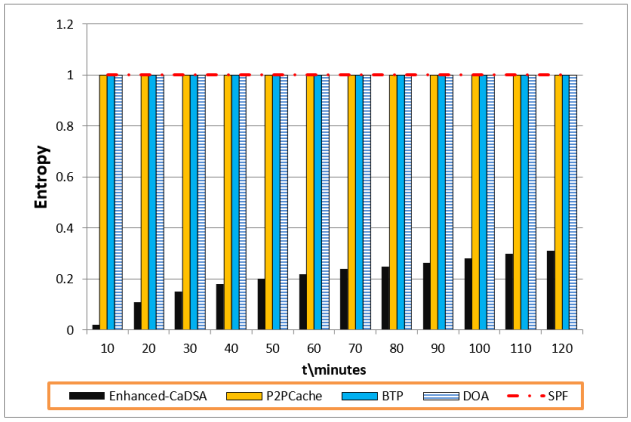

FIGURE 11. A Demonstration of the SPF Entropy.

After conducting our tests for all queries, we collected all the results and generated the figures. Then we checked the database to find out as to what data was recorded by the SP. Then we linked the true information about each user with their ID, and compared it with what was recorded by the SP.

We shall not compare the SPF method with the Obfuscation approach, because of the fundamental difference in the nature of applications. Our comparison focuses on the Privacy, and Performance Metrics [16], [33], [34], [36], [45], [51], [53]. As described earlier, the Privacy Metric consists of

1) K-Anonymity (the percentage of the real queries of the user, to the K (queries) sent to the SP)
2) Entropy (the amount of true data out of the whole data received by the SP from the same user
3) Ubiquity (degree of the spread of the user in the study area)

B. RESULTS OF SIMULATIONS

Figure 9 shows the efficiency of the SPF method over the other approaches according to the performance metric (Average time for response vs. queries number). This result reflects the use of one query for each request instead of a set as in the Dummy approach. This superiority is due to the fact that the SPF doesn't require any change or additional process for query processing. The improvement is also due to the fact that the SPF uses a pair of fogs and a pair of caches instead of just one, as is the case with other approaches, and also because it does not use any encryption, as is the case with the BTP Approach.

Figure 10 highlights the fact that the SPF sends fewer queries to the SP compared to other approaches, which is the result of using a pair of caches. This means, the cost of the SPF is the lowest. We have also accounted for the Cache-hit Ratio, which is the same in all approaches except Enhanced-CaDSA where it is the worst because the BTP, P2PCache, the SPF, and the DOA approaches send only one query each to the SP, and use the same method to manage the cache.

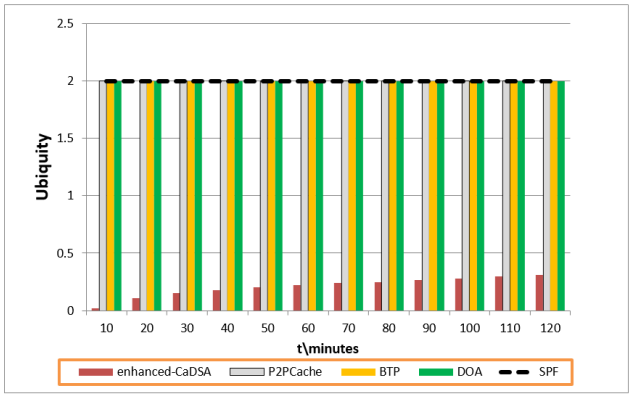

FIGURE 12. Ubiquity comparison with the SPF.

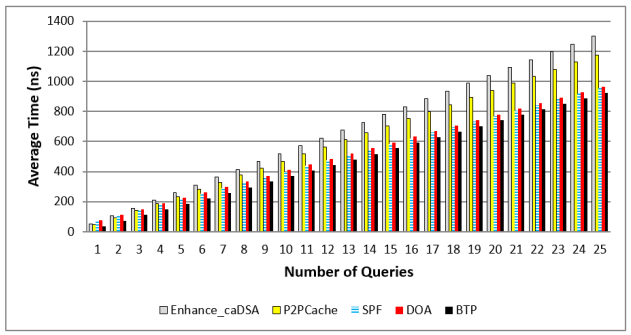

FIGURE 13. Average time Comparison with the SPF.

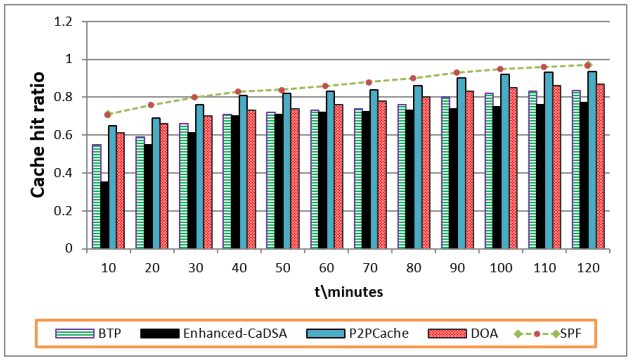

FIGURE 14. Cache Hit Ratio comparison with the SPF.

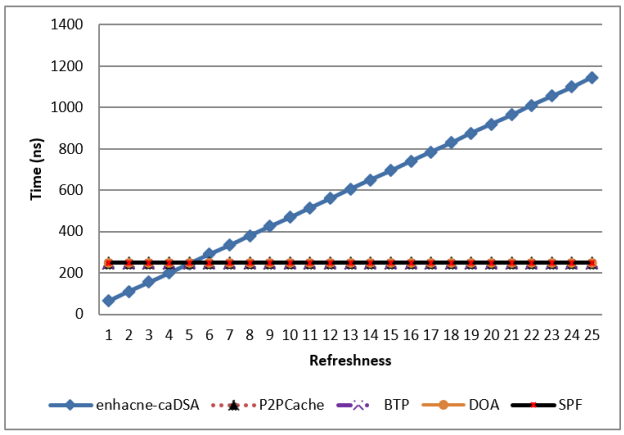

FIGURE 15. Comparison of refreshment with the SPF.

Figure 11 shows that the SPF method achieves the maximum amount of Entropy (E = 1), which is the same as was in the case of BTP, P2PCache, and DOA because the user in all of these approaches does not deal with the SP for query resolutions. However, the SPF method creates a higher level of privacy for location than P2PCache because P_1 and P_3 reside in different cells. Unlike in the case of DOA, the SPF approach does not affect the accuracy of results because it does not add noise or obfuscation to the query, and it does not deal with the same peers (TP) for different queries as is the case with the BTP.

As shown in Figure 12, the SPF achieves maximum ubiquity (U = 2), because this value is related to E value. Also, SPF achieves higher ubiquity because the users will be distributed randomly in a larger space compared to other approaches.

Figure 13 shows the average time required to carry out a search in cache. The BTP approach shows better performance because it used only one cache, not two like in the SPF method. However, the result of the SPF method is very close to the best result because it has used a Bloom filter to avoid the miss-hit time, and each cache is managed by a fog node.

As shown in Figure 14, the SPF received the highest Cache-hit ratio, because each time it had to deal with two caches in two fogs, not just one as in the other methods. In fact, this value is very important because it affects the privacy and performance Metrics.

Figure 15 depicts that the SPF, the BTP, P2PCache, and the DOA achieve a fixed ratio for freshness items in the cache by time, because they use the same algorithm in addition to a Bloom filter.

XI. CONCLUSION

The foregoing discussion has described characteristics, properties, advantages, disadvantages, implementations, and applications of the SPF method. We have demonstrated that the swapping mechanism in the SPF method eliminates most of the issues and open problems of the existing methods. We have also pointed out that this method works on the trust of the surrogate peer, and is not suitable when there are not enough users. It is well known that the number of users associated with applications in Smart Street at any given time is usually high enough for the SPF method to be very effective. In summary, the SPF method will protect user's privacy from the SP more than any known method, without affecting the accuracy of the core service and without significant drawbacks.

REFERENCES

[1] A. Whitmore, A. Agarwal, and X. Da, "The Internet of Things—A survey of topics and trends," *Inf. Syst. Frontiers*, vol. 17, no. 2, pp. 261–274, 2015, doi: 10.1007/s10796-014-9489-2.

[2] L. Da Xu, W. He, and S. Li, "Internet of Things in industries: A survey," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2233–2243, Nov. 2014, doi: 10.1109/TII.2014.2300753.

[3] H. Arasteh, V. Hosseinnezhad, V. Loia, A. Tommasetti, O. Troisi, M. Shafie-khah, and P. Siano, "Iot-based smart cities: A survey," in *Proc. IEEE 16th Int. Conf. Environ. Electr. Eng. (EEEIC)*, Jun. 2016, pp. 1–6, doi: 10.1109/EEEIC.2016.7555867.

[4] M. Yamin, A. M. Basahel, and A. A. Abi Sen, "Managing crowds with wireless and mobile technologies," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–15, Aug. 2018, doi: 10.1155/2018/7361597.

[5] B. Rashid and M. H. Rehmani, "Applications of wireless sensor networks for urban areas: A survey," *J. Netw. Comput. Appl.*, vol. 60, pp. 192–219, Jan. 2016, doi: 10.1016/j.jnca.2015.09.008.

[6] H. Sun, M. Yin, W. Wei, J. Li, H. Wang, and X. Jin, "MEMS based energy harvesting for the Internet of Things: A survey," *Microsyst. Technol.*, vol. 24, no. 7, pp. 2853–2869, Jul. 2018, doi: 10.1007/s00542-018-3763-z.

[7] J. Rittinghouse and J. Ransome, *Cloud Computing: Implementation, Management, and Security*. Boca Raton, FL, USA: CRC Press, 2010.

[8] R. Wang, Y. Liu, P. Zhang, X. Li, and X. Kang, "Edge and cloud collaborative entity recommendation method towards the IoT search," *Sensors*, vol. 20, no. 7, p. 1918, Mar. 2020, doi: 10.3390/s20071918.

[9] S. Oh, "Design of the smart application based on IoT," *J. Inst. Internet, Broadcasting Commun.*, vol. 17, no. 5, pp. 151–155, 2017, doi: 10.7236/JIIBC.2017.17.5.151.

[10] Y. Qu, M. R. Nosouhi, L. Cui, and S. Yu, "Privacy preservation in smart cities," in *Smart Cities Cybersecurity Privacy*. Amsterdam, The Netherlands: Elsevier, 2019., pp. 75–88, doi: 10.1016/B978-0-12-815032-0.00006-8.

[11] M. Ammar, G. Russello, and B. Crispo, "Internet of Things: A survey on the security of IoT frameworks," *J. Inf. Secur. Appl.*, vol. 38, pp. 8–27, Feb. 2018, doi: 10.1016/j.jisa.2017.11.002.

[12] S. M. P and V. Dr. D, "A study of data storage security issues in cloud computing," *Bonfring Int. J. Softw. Eng. Soft Comput.*, vol. 9, no. 2, pp. 05–07, Apr. 2019.

[13] K. Kolodziej and J. Hjelm, *Local Positioning Systems: LBS Applications and Services*. Boca Raton, FL, USA: CRC Press, 2006.

[14] I. Okta Sari, L. Andretti Abdillah, and K. Rizky Nova Wardhani, "Application location based service (LBS) location search palembang nature-based android," 2016, *arXiv:1602.06871*. [Online]. Available: http://arxiv.org/abs/1602.06871

[15] J. Zhou, Z. Cao, X. Dong, and A. V. Vasilakos, "Security and privacy for cloud-based IoT: Challenges," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 26–33, Jan. 2017, doi: 10.1109/MCOM.2017.1600363CM.

[16] A. Sen, F. A. Eassa, K. Jambi, M. Yamin, "Preserving privacy in Internet of Things: A survey," *Int. J. Inf. Technol.*, vol. 10, pp. 189–200, Dec. 2018, doi: 10.1007/s41870-018-0113-4.

[17] Y. Yang, L. Wu, G. Yin, L. Li, and H. Zhao, "A survey on security and privacy issues in Internet-of-Things," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1250–1258, Oct. 2017, doi: 10.1109/JIOT.2017.2694844.

[18] M. Dabbagh and A. Rayes, "Internet of Things security and privacy," in *Internet of Things From Hype to Reality*. Cham, Switzerland: Springer, 2019, pp. 211–238, doi:1 0.1007/978-3-319-99516-8_8.

[19] M. C. Kus Khalilov and A. Levi, "A survey on anonymity and privacy in bitcoin-like digital cash systems," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2543–2585, 3rd Quart., 2018, doi: 10.1109/COMST.2018.2818623.

[20] J. H. Ziegeldorf, O. G. Morchon, and K. Wehrle, "Privacy in the Internet of Things: Threats and challenges," *Secur. Commun. Netw.*, vol. 7, no. 12, pp. 2728–2742, Dec. 2014, doi: 10.1002/sec.795.

[21] G. Davis, "2020: Life with 50 billion connected devices," in *Proc. IEEE Int. Conf. Consumer Electron. (ICCE)*, Jan. 2018, p. 1, doi: 10.1109/ICCE.2018.8326056.

[22] A. P. G. Lopes and P. R. L. Gondim, "Mutual authentication protocol for D2D communications in a cloud-based E-Health system," *Sensors*, vol. 20, no. 7, p. 2072, Apr. 2020, doi: 10.3390/s20072072.

[23] A. A. Mutlag, M. Khanapi Abd Ghani, M. A. Mohammed, M. S. Maashi, O. Mohd, S. A. Mostafa, K. H. Abdulkareem, G. Marques, and I. de la Torre Díez, "MAFC: Multi-agent fog computing model for healthcare critical tasks management," *Sensors*, vol. 20, no. 7, p. 1853, Mar. 2020, doi: 10.3390/s20071853.

[24] Y. Jararweh, A. Doulat, O. AlQudah, E. Ahmed, M. Al-Ayyoub, and E. Benkhelifa, "The future of mobile cloud computing: Integrating cloudlets and mobile edge computing," in *Proc. 23rd Int. Conf. Telecommun. (ICT)*, May 2016, pp. 1–5, doi: 10.1109/ICT.2016.7500486.

[25] R. Mahmud, R. Kotagiri, and R. Buyya, "Fog computing: A taxonomy, survey and future directions," in *Internet Everything*. Singapore: Springer, 2018, pp. 103–130.

[26] F. Song, Z.-Y. Ai, J.-J. Li, G. Pau, M. Collotta, I. You, and H.-K. Zhang, "Smart collaborative caching for information-centric IoT in fog computing," *Sensors*, vol. 17, no. 11, p. 2512, Nov. 2017, doi: 10.3390/s17112512.

[27] V. Kumar, A. Laghari, S. Karim, M. Shakir, and A. Brohi, "Comparison of fog computing & cloud computing," *Int. J. Math. Sci. Comput*, vol. 1, pp. 31–41, Jan. 2019, doi: 10.5815/ijmsc.2019.01.03.

[28] R. Kumar Naha, S. Garg, and A. Chan, "Fog computing architecture: Survey and challenges," 2018, *arXiv:1811.09047*. [Online]. Available: http://arxiv.org/abs/1811.09047

[29] R. Deng, R. Lu, C. Lai, T. H. Luan, and H. Liang, "Optimal workload allocation in fog-cloud computing toward balanced delay and power consumption," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1171–1181, Dec. 2016, doi: 10.1109/JIOT.2016.2565516.

[30] H. F. G. B. Atlam Wills, "IoT security, privacy, safety and ethics," in *Digital Twin Technologies and Smart Cities Internet of Things*, M. Farsi, A. Daneshkhah, A. Hosseinian-Far, and H. Jahankhani, Eds. Cham, Switzerland: Springer, 2020.

[31] A. A. A. Sen, F. B. Eassa, M. Yamin, and K. Jambi, "Double cache approach with wireless technology for preserving user privacy," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–11, Aug. 2018, doi: 10.1155/2018/4607464.

[32] Z. B. Celik, E. Fernandes, E. Pauley, G. Tan, and P. McDaniel, "Program analysis of commodity IoT applications for security and privacy: Challenges and opportunities," *ACM Comput. Surv.*, vol. 52, no. 4, pp. 1–30, Sep. 2019.

[33] M. Yamin, Y. Alsaawy, A. B. Alkhodre, and A. A. Abi Sen, "An innovative method for preserving privacy in Internet of Things," *Sensors*, vol. 19, no. 15, p. 3355, Jul. 2019, doi: 10.3390/s19153355.

[34] S. S. Albouq, A. A. A. Sen, A. Namoun, N. M. Bahbouh, A. B. Alkhodre, and A. Alshanqiti, "A double obfuscation approach for protecting the privacy of IoT location based applications," *IEEE Access*, vol. 8, pp. 129415–129431, 2020, doi: 10.1109/ACCESS.2020.3009200.

[35] S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini, "Security, privacy and trust in Internet of Things: The road ahead," *Comput. Netw.*, vol. 76, pp. 146–164, Jan. 2015, doi: 10.1016/j.comnet.2014.11.008.

[36] B. Niu, Q. Li, X. Zhu, G. Cao, and H. Li, "Enhancing privacy through caching in location-based services," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2015, pp. 1017–1025, doi: 10.1109/INFOCOM.2015.7218474.

[37] J.-N. Luo and M.-H. Yang, "Unchained cellular obfuscation areas for location privacy in continuous location-based service queries," *Wireless Commun. Mobile Comput.*, vol. 2017, pp. 1–15, Dec. 2017, doi: 10.1155/2017/7391982.

[38] A. D. Lahe and P. Kulkarni, "Location privacy preserving using semi-TTP server for LBS users," in *Proc. 2nd IEEE Int. Conf. Recent Trends Electron., Inf. Commun. Technol. (RTEICT)*, May 2017, pp. 605–610, doi: 10.1109/RTEICT.2017.8256668.

[39] S. Kadhe, B. Garcia, A. Heidarzadeh, S. El Rouayheb, and A. Sprintson, "Private information retrieval with side information," in *Proc. 55th Annu. Allerton Conf. Commun., Control, Comput.*, Oct. 2019, pp. 1–4, doi: 10.1109/ALLERTON.2017.8262860.

[40] T. Peng, Q. Liu, D. Meng, and G. Wang, "Collaborative trajectory privacy preserving scheme in location-based services," *Inf. Sci.*, vol. 387, pp. 165–179, May 2017, doi: 10.1016/j.ins.2016.08.010.

[41] X. Zhu, H. Chi, B. Niu, W. Zhang, Z. Li, and H. Li, "MobiCache: When k-anonymity meets cache," in *Proc. IEEE Global Commun. Conf.*, Dec. 2013, pp. 820–825, doi: 10.1109/GLOCOM.2013.6831174.

[42] Tankard, C, "What the GDPR means for businesses," *Netw. Secur.*, vol. 6, no. 1, pp. 5–8, 2016, doi: 10.1016/S1353-4858(16)30056-3.

[43] C. Bartolini, S. Daoudagh, G. Lenzini, and E. Marchetti, "GDPR-based user stories in the access control perspective," in *Proc. Int. Conf. Qual. Inf. Commun. Technol.* Cham, Switzerland: Springer, 2016, pp. 3–17, doi: 10.1007/978-3-030-29238-6_1.

[44] S. Wachter, "Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR," *SSRN Electron. J.*, vol. 7, pp. 436–449, Dec. 2018, doi: 10.2139/ssrn.3083554.

[45] M. Yamin and A. A. A. Sen, "Improving privacy and security of user data in location based services," *Int. J. Ambient Comput. Intell.*, vol. 9, no. 1, pp. 19–42, Jan. 2018, doi: 10.4018/IJACI.2018010102.

[46] S. Hayashida, D. Amagata, T. Hara, and X. Xie, "Dummy generation based on user-movement estimation for location privacy protection," *IEEE Access*, vol. 6, pp. 22958–22969, 2018, doi: 10.1109/ACCESS.2018.2829898.
[47] Z. Hu, S. Havrylov, I. Titov, and S. B. Cohen, "Obfuscation for privacy-preserving syntactic parsing," 2019, *arXiv:1904.09585*. [Online]. Available: http://arxiv.org/abs/1904.09585
[48] Y. Alsaawy, B. Alkhodre, A. Sen, and S. M. Siddiqui, "Swap obfuscation technique for preserving privacy of LBS," *Int. J. Academic Sci. Res.*, vol. 7, no. 2, pp. 11–19, 2019.
[49] G. Sun, S. Cai, H. Yu, S. Maharjan, V. Chang, X. Du, and M. Guizani, "Location privacy preservation for mobile users in location-based services," *IEEE Access*, vol. 7, pp. 87425–87438, 2019, doi: 10.1109/ACCESS.2019.2925571.
[50] J. H. Mun and H. Lim, "Cuckoo Bloom filter," in *Proc. Int. Conf. Electron., Inf., Commun. (ICEIC)*, Jan. 2019, pp. 1–2, doi: 10.23919/ELINFOCOM.2019.8706343.
[51] K. G. Shin, X. Ju, Z. Chen, and X. Hu, "Privacy protection for users of location-based services," *IEEE Wireless Commun.*, vol. 19, no. 1, pp. 30–39, Feb. 2012, doi: 10.1109/MWC.2012.6155874.
[52] M. Wernke, P. Skvortsov, F. Därr, and K. Rothermel, "A classification of location privacy attacks and approaches," *Pers. Ubiquitous Comput.*, vol. 18, no. 1, pp. 163–175, Jan. 2014.
[53] M. Alrahhal, M. Khemakhem, and K. Jambi, K, "A survey on privacy of location-based services: Classification, inference attacks, and challenges," *J. Theor. Appl. Inf. Technol.*, vol. 95, p. 24, May 2017.
[54] P. Dheena, G. Raj, G. Dutt, and S. Jinny, "IoT based smart street light management system," in *Proc. IEEE Int. Conf. Circuits Syst. (ICCS)*, Dec. 2017, pp. 368–371, doi: 10.1109/ICCS1.2017.8326023.
[55] G. Soumyalatha, "Study of IoT: Understanding IoT architecture, applications, issues and challenges," in *Proc. 1st Int. Conf. Innov. Comput. Netw.*, 2010, pp. 1–5.

MOHAMMAD YAMIN received the Ph.D. degree in mathematics from The Australian National University (ANU), Canberra, Australia, in 1983. He is currently a Professor of management information systems (MIS) with the Faculty of Economics and Administration, King Abdulaziz University, Jeddah, Saudi Arabia. He is also an Honorary Level D Academic at the Research School of Computer Science, ANU. He has also served at the University of Canberra for 11 years and the Department of Human Services of the Australian Federal Government. His research interests include privacy of users in the IoT applications, wireless sensor networks, crowd management and prevention of stampedes, and health informatics. He has published about 100 research articles in journals and proceedings, most of which are indexed in the Web of Science and Scopus database. He is also an Editor of the *International Journal of Information Technology* (Springer), indexed in Scopus Database.

ADNAN AHMED ABI SEN received the Ph.D. degree in computer sciences from King Abdulaziz University, Jeddah, Saudi Arabia. In his academic career, he has supervised dozens of graduation projects in the different fields of computer science. He is currently a Business and Systems Analyst with Islamic University, Saudi Arabia, and a Lecturer for the Network Security Diploma Program. He is an experienced and established Researcher in privacy and mobile computing. He also has expertise in systems analysis and development. He has published about 20 research articles, many of which are indexed in Thomson and Reuters database.