IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 14, 2020, accepted July 6, 2020, date of publication July 14, 2020, date of current version July 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3009200

A Double Obfuscation Approach for Protecting the Privacy of IoT Location Based Applications

SAMI SAAD ALBOUQ, ADNAN AHMED ABI SEN, ABDALLAH NAMOUN, NOUR MAHMOUD BAHBOUH, AHMAD B. ALKHODRE, AND ABDULLAH ALSHANQITI

Faculty of Computer and Information Systems, Islamic University, Al-Madinah 42351, Saudi Arabia

Corresponding author: Adnan Ahmed Abi Sen (adnanmnm@iu.edu.sa; adnanmnm@hotmail.com)

This work was supported by the Deanship of Research at Islamic University of Madinah (IUM) for application-based research under Project 101.

ABSTRACT Connected vehicles and smart cars have become highly reliant on location-based services (i.e. LBS) to provide accurate, personalized and intelligent services. However, location-based services have endangered its users to considerable risks concerning the privacy and security of users' personal data. Although existing research provides a myriad of methods to improve and protect user privacy in LBS applications, most of these methods are concerned with handling static queries and non-mobile objects only. Moreover, various issues and challenges still persist with regards to the need to trust third parties, overloading of the user, and low accuracy of the returned results. This paper contributes a Double Obfuscation Approach (referred to as DOA) that applies two phases of obfuscation consecutively whilst integrating two differing privacy protection approaches, namely Obfuscation and Trusted Third Party, and two techniques, namely fog caching technology and mix zone. In essence, the DOA obfuscates and hides the identity and location of its users using the fog nodes, which operate as a trusted third party (TTP), and without the need to reveal the identity of the users or trust the cooperating nodes. Moreover, this paper presents a DOA algorithm that improves the overall user privacy and system performance using the fog nodes, which split the responses of each query into five parts, thus reducing the processing time of the results by the user and enhancing the overall accuracy where the user directly selects the most suitable parts based on his current location. Overall, the hybrid DOA approach empowers the users of connected vehicle applications to protect their privacy through an algorithm that caters for the dynamic nature of user queries and mobility of objects. The results of our comparative simulations against well-known hybrid privacy protection methods demonstrate the superiority of the proposed Double Obfuscation Approach especially with respect to user privacy whilst maintaining a nominal overhead on the user, reduced response time and high accuracy of the obtained results.

INDEX TERMS Obfuscation, trusted third party, location based services, Internet of Things, privacy, connected vehicles.

I. INTRODUCTION

The Internet of Things (i.e. IoT) has changed the way people use services in all fields of life, such as health, transportation, business, education, energy, communication, entertainment, among others [1], [2]. IoT integrate digital information and systems into the real world [3]. In fact, everything around us is now connected to the Internet and is empowered to sense, process, and share data to serve various human needs. The world is full of smart objects that act as constant observers of the things we perform daily [4]. Since most of these objects, e.g. wireless network sensors (WSNs) and radio identifiers (RFID), are limited with respect to the computing power and memory they can hold [5], [6], cloud computing capabilities can be exploited to store and process the data of IoT applications [7], [8].

Connected vehicles represent an ideal example of the most common applications of smart cities and intelligent transportation systems today [9], [10]. Connected vehicles aim to ameliorate the existing commuting services by reducing traffic congestion and accidents, speeding up access to the required places or points of interest (i.e. POI), and supporting health and emergency applications [11], [12]. In principle, each smart vehicle is connected to a service provider (i.e. SP) where its location is periodically sent along with its query and/or destination. The connected vehicles may also

The associate editor coordinating the review of this manuscript and approving it for publication was Ana Lucila Sandoval Orozco.

communicate to each other or cooperate wirelessly to achieve a particular goal [13].

Despite their benefits, smart objects have become a constant observer of users' lives and may expose, intentionally or otherwise, their private data, which are stored in the cloud as a result of penetration by malicious entities. In some instances, the SP may be part of the privacy breach. For example, collecting and analyzing the stored places and times of users' presence on a regular basis could help in inferring various important information about the users and recognizing their behaviors, social habits, customs, hobbies, workplaces, times of travel, and health status [14], [15].

The privacy and security of data have been for long the main concern of many researchers and developers of software applications, including those developed using modern technologies [16], [17]. This interest has expanded to reach governments, which have established strict policies and laws to restrain the access of companies and service providers to users' sensitive data in a bid to protect citizens' privacy [18].

Many techniques have emerged in the area of privacy protection, such as the Dummy [30], K-Anonymity [17] and [31], TTP [31], Obfuscation [32], PIR [33], and Clacking Area [37] techniques. Unfortunately, these techniques still suffer from serious issues, such as the need to trust a third party, the weak accuracy of the results, the overload on the user, as well as the adverse effects on the performance of the whole system [19], [20]. These techniques are discussed in detail in the related works section.

Our research presents a Double Obfuscation Approach for preserving user privacy by addressing the shortcomings of privacy protection approaches (e.g. the need to trust a third party and overloading of the user) by integrating the advantages of two privacy techniques and two modern technologies (i.e. fog computing and caching). Fog computing has resolved many problems of cloud computing, especially in terms of increasing availability and mobility, supporting latency-sensitive applications, reducing the load, and improving the overall system performance by being on the edge of the network and close to the user [21]. Moreover, fog computing can process data on behalf of the cloud before sending it to the service provider [22]. It also enables collaboration between the fog nodes by exploiting their caches, thus providing additional features like improving the cache-hit ratio [23].

Figure 1 shows the hierarchical relationship between the Internet of Things apps and cloud and fog computing. The bottom layer represents a set of mobile IoT objects that provide access to raw data, which are collected and pre-processed by the fog nodes. Next, the core fog layer organizes the work of the nodes and provides the necessary extra resources to process the collected data. The cloud computing layer is used for the central processing and permanent storage of data. Finally, the IoT applications in the top layer act as service providers that fulfill user needs and queries [24], [25].

This research makes several interesting contributions as highlighted by the following points:

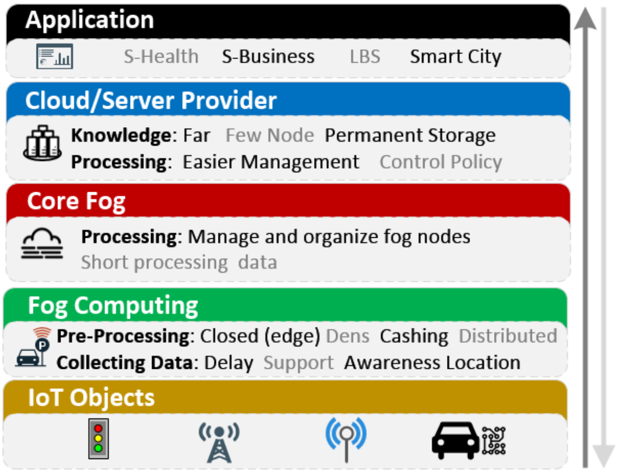

Application
S-Health
S-Business
LBS
Smart City
Cloud/Server Provider
Knowledge: Far Few Node Permanent Storage
Processing: Easier Management Control Policy
Core Fog
Processing: Manage and organize fog nodes
Short processing data
Fog Computing
Pre-Processing: Closed (edge) Dens Cashing Distributed
Collecting Data: Delay Support Awareness Location
IoT Objects

FIGURE 1. **An overview of the IoT architecture and its applications.**

1. Introduce a new approach, named Double Obfuscation Approach (DOA), to protect user privacy (e.g. identity, query, and location) in dynamic environments through the integration of Obfuscation and Trusted Third Party techniques.
2. Increase the level of user privacy compared to using any of previous techniques separately, i.e. Obfuscation or Trusted Third Party.
3. Reduce the load on the user (e.g. through the preparation and processing of queries) and improve the system performance (e.g. the turnaround time taken for sending the query and receiving the response) compared to the traditional Obfuscation or Dummy approach [19], [20].
4. Eliminate the reliance on a trusted third party compared to the traditional TTP approach.
5. Improve the accuracy of the returned results compared to the existing Obfuscation methods.
6. Exploit the advantages of fog computing and its caching technology to improve the response time of the system.
7. Improve the system resistance against other types of attacks [26], such as the malicious TP or correlation attack . . . etc., in comparison to the traditional obfuscation and TTP approach 19], [20].

The remainder of this paper is organized into seven sections. Section two reviews the latest privacy protection techniques highlighting their merits and weaknesses. Section three explains the proposed Double Obfuscation Approach along and its underlying concepts. Section four details the steps of the DOA algorithm. Section five provides an example for the potential application of the DOA. Section six discusses several metrics that were used for comparing and evaluating the results of various privacy techniques. Section seven highlights the findings of this research and suggests key implications for the privacy of mobile IoT objects.

II. LITERATURE REVIEW

This section presents the research works that relate to the famous techniques used to protect the privacy of users in IoT, followed by an overview of the connected vehicles and their real-life applications.

Privacy refers to the individual's right to decide who, when and how their personal data can be reached and/or used. Data privacy, however, differs from data security as it refers to protecting the confidentiality of the data being exchanged between mutually trusted parties, as well as preserving the integrity of these data and preventing its modification, and finally ensuring the availability of services between the involved parties. This can be achieved by preventing the hacking of these parties by external attackers. Data privacy ensures that data are not accessed without permission or used to infer any related information (such as gender, age group, habits . . . etc.).

Data protection refers to the blocking of external attacks by service providers or malicious parties from revealing or profiling the identity of the users. Although different in definition, privacy and security must be integrated within any system to a certain degree [16], [19], [20], and [27]. In data privacy, the data may not be confidential since collecting a considerable amount of data and analyzing it may expose unexpected private information. Additionally, the second party to which the data is being sent may be malicious. For example, a malicious SP can collect user data like her locations on specific times and then analyze these data to infer private information that are not related to the supplied services. Hence, privacy techniques aim to protect the privacy of users of services and IoT applications without making their data vulnerable to unwarranted exploitation. In other words, privacy focuses on preventing the identification, traceability and profiling of users [28].

A. PRIVACY PROTECTION APPROACHES

Protecting privacy in location-based services is one of the most important sought-after goals by researchers. Several approaches and methods have been proposed to maintain user privacy within applications that utilize location-based services, such as those used in connected vehicles. However, these approaches still suffer from various open issues. Table 1 lists the major techniques used for data protection and highlights their advantages and disadvantages [19], [20], [28]. Moreover, Table 1 identifies the exact privacy enhancements offered by each privacy technique.

B. HYBRID PRIVACY PROTECTION APPROACHES

Latest research efforts attempted to integrate multiple privacy approaches and technologies in a bid to improve the privacy performance, such as the integration of the cache technology with the TTP [36]. Below we discuss the main hybrid privacy protection approaches.

The Enhanced Cache technique combines the distributed cache and dummy queries. However, this method overloads the user during the creation of the dummies and does not manage the cache independently. Moreover, the user is still somewhat connected to the service provider directly and sends real information, which subjects his privacy to potential vulnerabilities [38]. The Preserving Privacy Cyber Services (PPCS) approach generates dummies that are difficult to discover by the SP [48]; however, it still suffers from similar issues to the Enhance Cache techniques [38].

The Peer-to-Peer Cache (P2PCache) approach uses the cache to improve the performance and relies on the collaboration between the nodes to generate real dummies [39]. Queries are first exchanged between the users and then forwarded to the designated service provider. Although this approach conceals the real identity of the users from the service provider, it reveals information about the location of the query since the collaborating users are neighbors. Moreover, the collaboration is highly dependent on the level of trust between the users and their availability.

The CAST approach [34] exploits the cache available within the peers to reduce the connection to the SP. In principle, the peer would request answers to their queries from their peers first. Similarly, Hiding in Crowd approach uses the same concept [50]. Both approaches suffer from two main disadvantages; firstly, they do not implement a clear way of managing the peers and the way they communicate with each other; secondly, they both rely on trust between the peers themselves. Although the CAST approach attempted to quantify the quality and reputation of each peer based on the percentage of correct responses it provides and the level of interaction, this does not guarantee that the cooperating peer will not breach the privacy of other peers.

The Double Cache Approach (DCA) uses two parts of the cache within the access point; the first part exchanges the queries between the peers whilst the second part stores the results of these queries [40]. This approach attempts to overcome the trust factor between the peers by shifting it to the access point. However, managing and protecting the access points remain an open issue.

To overcome the above challenges persisting within the existing privacy approaches [19], [20], this research proposes a novel method, called Double Obfuscation Approach (DOA), that enhances user privacy by utilizing the resources of the fog nodes to serve the IoT applications [41]. In essence, the Double Obfuscation Approach (DOA) integrates the Obfuscation and TTP approaches, with other techniques namely fog cache and mix-zone to achieve better protection for the users than when each approach is used separately. Our approach is then validated against three modern hybrid techniques, which are used as a comparison benchmark, namely the Enhanced Cache [38], P2PCache [39], and DCA Double Cache [40].

III. THE CONCEPT OF DOUBLE OBFUSCATION APPROACH

A. OVERVIEW OF THE DOA

Broadly speaking, the Double Obfuscation Approach (aka DOA) integrates four advantageous privacy protection practices, where two are the primary privacy protection approaches (i.e. Obfuscation [44] and TTP [31]) and the remaining two are complementary techniques (i.e. Cache [40]

TABLE 1. A summary comparison of the major privacy protection techniques.

Name of Privacy Protection Technique	Way of Working	Key Advantages	Disadvantages	Privacy Enhancement by this Approach
Privacy Policies	SP has to announce clearly how it will use the client's data to ensure users' awareness	Prevents SPs from exploiting client information	There is no guarantee that the SP or some nodes will not be malicious	Launch of various government initiatives to support this approach [19]
Anonymity [29]	Uses an alias or hashing code instead of the real identity	Protects user identification	Not effective if the service provider is the attacker itself	Forcing the user to change his identity (i.e. nickname) periodically in each new area that he visits. This is called Mix-Zone
Mix-Zone [49]	Creates a mix area through cooperation among K users or by depending on a TTP to refine their pseudonyms together to prevent association with the old ones	Protects the identity of users and avoids tracking attacks	Ineffective if there is not enough users in the selected area	Finding a method for generating a suitable mix zone if the density of the users is low
Dummy [30]	Sends multiple queries with false locations in addition to the real query as one set in each connection to the SP	Does not require trust in another party and protects the real location / type of query	Inappropriate in dynamic environments where the user repeatedly sends queries. In this case the real query and false queries can be detected and noticed by an attacker. Moreover, the dummy causes additional overload on the user and the system	Generating smarter dummies, which will be difficult for an attacker to distinguish from real queries, as well as improvements to the performance
Trusted Third Party (TTP) [31]	Relies on a trusted third party for privacy protection. TTP hides the user's identity by communicating with SP on behalf of the user	Protects user privacy from the main SP	Needs to trust a third party; the privacy of the user is threatened if this third party is malicious	Increasing the level of privacy by hiding the user identity among many other users, which is known as K-Anonymity. As well as finding a way to reduce the level of trust required when dealing with TTP
Obfuscation [32]	Distorts the data before sending or adding noise to it. For example, sending a location that is close to the real one, or sending an area "clacking area" instead of a specific location	Protects the privacy of a location in dynamic environments; Does not require trust in another party	Causes adverse effects on the accuracy of the returned results and on the overall performance of the system	Protecting privacy without significantly affecting the accuracy of results and system performance
Private Information Retrieval (PIR) [33]	Withdraws a large amount of data from the SP, so SP cannot determine which real part of the data is wanted	Preserves the privacy without the need for a third-trusted party; in addition, it relies on encryption to increase security	Affects the performance of the whole system adversely, so it cannot be applied by a user or a device with normal resources	Working on improving the performance of these systems such as finding lightweight encryption methods
Peers Cooperation [34]	Users collaborate to increase their privacy by sharing the results of the queries or data between them	Improves the privacy of all peers, and prevents an attacker to distinguish between them	Needs to trust another peer and impacts the performance negatively due to the communication between peers	Finding ways to facilitate user collaboration and making it more effective, and avoiding the threat of trusted peers
Caching [35]	Stores some of the results of the previous queries for future requests in order to reduce communication with the SP	Improves the privacy and performance due to the reduction in the number of connections to the SP	The position and management of the cache itself is still an open issue	Improving the cache-hit ratio by storing queries to serve potential future requests, and reducing the search miss-time to enhance the performance
Cloaking Area [19, 37]	Protects the physical location of the user in the selected area by cooperation or by relying on the TTP. Users agree to send same location (i.e. location of selected area) to the SP.	Can be used to protect the privacy of location by Obfuscation or protecting the privacy of query by sending all users' queries together	Similar disadvantages to the Peers Cooperation, Obfuscation and TTP approaches	Various new methods for creating a cloak area exist such as, Centre cloak, Kasper, Interval & Special, Quad Tree, Hilbert, etc. [19]
K-Anonymity [31]	Hides the user ID among K other users by using on an Anonymizer (i.e. TTP)	Protects the identity and location privacy of users	Needs to trust a third party	Finding methods avoid full trust in the Anonymizer
Hybrid Modern Approaches	Relies on the integration of more than one technique to create an enhanced protection approach	Achieves more advantages compared to the use of a single technique	Complexity and overhead	Many new techniques including CAST, PPCS, P2Pcache, DCA, BTP, etc. Further details of these approaches are discussed in the subsequent section [34, 48, 39, 40, 25]

and Mix-zone [49]) to avoid the weaknesses of each approach when used separately and boost the system protection and performance.

In the DOA architecture (see Figure 2), the users do not communicate directly with the SP but rather talks to the closest fog node (step 1 and step 4), which normally represents the local TTP that is available in the current cell. In the DOA, the fog node operates on behalf of the user and thus conceals their identity from the SP (step 2 and step 3). Moreover, our architecture applies the concept mix-zone, where each user is assigned a new randomly generated nickname upon visiting a new cell to make it harder for the fog nodes to track and follow the new locations of the users. The DOA architecture applies two phases of obfuscation (i.e. thus the name double obfuscation) to protect the privacy of user's location from both the fog node and the SP. Finally, the DOA architecture capitalizes on the cache of the fog to respond to future queries and thereby enhance the privacy and performance.

It is worthwhile to note that the fog nodes play the role of the TTP. In fact, it is an improved TTP because the user does not have to fully trust the nearest fog node. In that sense, each fog node represents a TTP within the cell. Therefore, our proposed architecture creates a distributed TTP which is hosted within the fog nodes of the area. This eliminates the need to have one centralized TTP, thus improving the privacy of the users and the performance of the system. This way the system eliminates the issue of a single point of failure and improves its services.

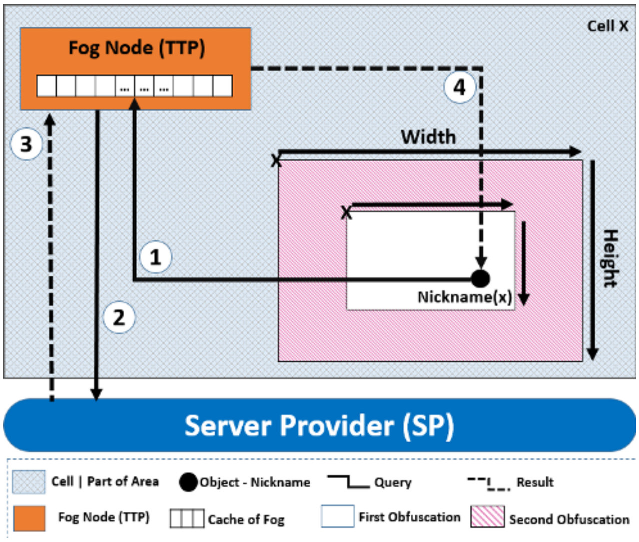


FIGURE 2. The proposed double obfuscation approach, numbers represent execution flow of requests.

B. KEY CONCEPTS OF THE DOA

The important concepts pertaining to our DOA architecture (Fig 2) are explained below.

1. **Obfuscation Approach**: obfuscation is used to protect the exact location of the user by adding noise or encapsulating his real location inside the selected area. The DOA algorithm selects a small area as the first obfuscation area (represented using the white rectangle in Figure 2). This phase protects the privacy of user's location from the fog node, which could be malicious. This is due to the fact that the user will send his query to the fog node at the end of this phase.
2. **TTP Approach**: the DOA divides the area into a number of cells (N), where each cell will be managed by a dedicated fog node. The fog node represents the TTP of this cell only. The fog node provides important functions in our system including:
   a. It protects the user location by generating the second obfuscation area (represented by the pink rectangle in Figure 2) around the first obfuscation area to make it bigger and add an extra layer of protection to the real location of the user.
   b. It conceals the user identity and his queries by sending each query to the SP on behalf of the user.
   c. It processes the returned result by the SP by dividing it into five parts to reduce the overhead of the obfuscation technique on the user's device.
   d. It takes advantage of the cache of the fog node in order to answer future user queries without requiring to connect to the SP. This will enhance both the privacy and performance of system.
3. **Caching Technique**: caching is used to save some of the users' queries and answers to help in responding to future queries without connecting to the SP again. The DOA uses the cache of fog nodes to achieve that, where each fog node manages and protects its cache independently.
4. **Mix-Zone Technique**: the DOA employs a simple concept, called mix-zone, where it generates a new nickname for the user/object as soon as they enter a new cell and deals with a new fog node (TTP). Each cell in the area is considered as 'a mix-area' for any K users wo enter this cell.

C. MAIN STEPS OF THE DOA

Below we present the main steps of the DOA algorithm and link them to the pseudo code.

1. Firstly, the region is divided into several cells and each cell is managed by an anonymized fog node in a similar way to the traditional TTP. However, the fog node is responsible for only a small area (i.e. the designated cell). Therefore, the reliance on the fog computing by users (i.e. smart vehicles) will be limited and only during the presence of the vehicle inside the region. This way reduces the risks of successful attacks from the malicious fog nodes as opposed to the traditional TTP approach. The user will not send his exact location to the fog as explained in the below steps. Moreover, the user will use a different alias/nickname every time she connects to a new fog node in order to prevent the tracking of her path.
2. The obfuscation process consists of two phases as illustrated in Figure 3 (i.e. plain and dotted areas). In the

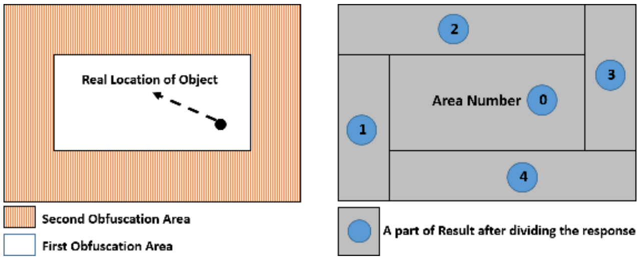


FIGURE 3. The double obfuscated phases, and divided areas of result.

first phase, the user selects a small area as a simple confusion area for his query's location (Step 7 in DOA algorithm) before sending it to the fog node in the cell (Step 8 in DOA algorithm). The user does not need to fully trust the fog node in this case. Advantageously, the load on the user will be reduced due to the small area of obfuscation and the user will not need a large area because she will deal with the fog node only for a limited time and within a specific region. The fog node, therefore, will be unable to track the path of the user movement. In the second phase, a second obfuscation is created around the first obfuscation area by the fog node on behalf of the user (Step 31 in DOA algorithm). This will increase the additional obfuscation in an unbalanced way around the perimeter of the first obfuscation zone, prior to sending the queries to the service provider.

3. The fog node will send the query, after the creation of the second obfuscation area, to the service provider on behalf of the user (Step 32 in DOA algorithm). Since the user does not communicate with the service provider directly, he will preserve his privacy from potential threats posed by the service provider.
4. The service provider will receive the data from the fog node without the identity and correct location of the user. Once the data are processed, the service provider will return the results to the fog node only (Step 32 in DOA algorithm).
5. After receiving the result, the fog node will divide it into five zones (Step 33 in DOA algorithm). The central zone will include the primary area requested by the user (i.e. the first obfuscation zone), surrounded by four other secondary zones as shown in Figure 3. The fog node then sends the new results to the end-user (i.e. the requester) (Step 34 in DOA algorithm).
6. The user receives the five zones of the result (Step 8 in DOA algorithm) and selects one or more of these zones according to his current position (from step 9 to step 26 in DOA algorithm). Thus, the accuracy of the results will be improved significantly compared to all previous obfuscation techniques. In addition, the load on the user will be lowered whilst his location is concealed from the fog node.
7. Note that as the user moves into a new zone, she will contact a new fog node. In doing so, the user will utilize a new alias to prevent the tracing to her path, even when there is more than one malicious fog node.
8. Finally, the DOA uses the cache of each fog node to store user queries for future requests. The cache will significantly reduce the number of connections to the SP, thus improving the overall performance and privacy of the system. Moreover, the DOA uses the Bloom Filter, which is a hash function that accelerates the search within the cache and avoids time wastage in the miss-hit cases.

IV. THE PROPOSED ALGORITHM OF THE DOA

The DOA algorithm takes as inputs the level of obfuscation (L) and user query (Q). Overall, as a user enters an obfuscation level (L) that he desires to use in the first phase, where L is an integer measured in meter and Q represents the user query or destination, the algorithm sends this query to the nearest fog node in the form Nickname_of_User (Q, X1, Y1, W1, H1), where W1 and H1 represent the width and height of the first obfuscation area respectively. Next, the fog node will search for the answer of the query in its cache. If the answer is not found, then the fog node will apply the second obfuscation phase and then send the user query to the SP in the form Fog_Id (Q, X∼, Y∼, W2, H2), where W2 and H2 represent the width and height of the second obfuscation area respectively.

The SP will search for the answer of Q inside the received area and return a set of answers in the following format, Answers = [[Answer1, Latitude1, Longitude1], [Answer2, Latitude2, Longitude2], [etc. . . .]], where the response represents a set of points of interest (i.e. POIs). Subsequently, the designated fog node will receive the results and filter the answers and divide them into five parts according to their locations. It will then return the set of results as [Part1, Part2, Part3, Part4, Part5] where each part is similar to the Answers area but surely has less results. The output of DOA algorithm is an array of answers containing the points of interest = {Part1[[Ans1,Lat1,Lon1], [Ans2,Lat2,Lon2], [. . .]], Part2[. . . .], etc.}.

More precisely, the DOA algorithm works by addressing two main ideas of the Double Obfuscation Approach as elaborated below.

Firstly, the DOA divides the obfuscation process into two subsequent phases (See Algorithm 1).

A. Phase One: the user creates the first obfuscation area (e.g. denoted as R1 in Figure 4) by selecting the level of obfuscation (L) where L is an integer number (measured in Meter). Generally, an obfuscation area can be a circle or a rectangle. If it is a circle then L will represent the radius and the current location of user (X,Y) is the center of the circle; However, after creating a circular obfuscation area, its center has to be shifted randomly to a new point (X∼, Y∼) where X∼, Y∼<=L.

In our DOA, we used a rectangular obfuscation area (Figure 4). After the user has entered the desired

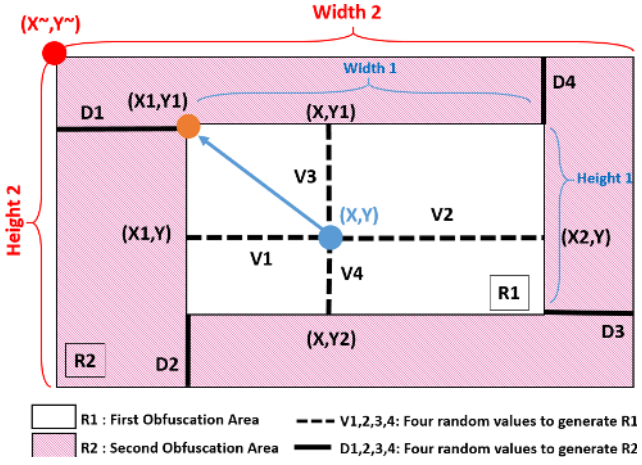

(X~,Y~)
Width 2
Width 1
D1
(X1,Y1)
(X,Y1)
D4
V3
Height 1
Height 2
(X1,Y)
(X,Y)
V2
(X2,Y)
V1
V4
R1
(X,Y2)
D3
R2
D2
R1 : First Obfuscation Area
R2 : Second Obfuscation Area
V1,2,3,4: Four random values to generate R1
D1,2,3,4: Four random values to generate R2

FIGURE 4. **Calculation of obfuscation areas (R1 and R2).**

obfuscation level L, the DOA algorithm performs the following steps:

- Generate four random variables V1, V2, V3, and V4 to create the first area, where all of these variables are assigned integer value that are <= L
- Calculate $X1 = X - V1$ & $X2 = X + V2$ & $Y1 = Y - V3$ & $Y2 = Y + V4$, where $width1 = X2 - X1$ & $height1 = Y2 - Y1$
- Generate the first obfuscation area (R1), which is represented by (X1, Y1, Width1, Height1)

B. Phase Two: fog computing will generate the second obfuscation area (denoted as R2) through the following steps:

- Generate random four variables D1, D2, D3, D4 to expand area R1 to area R2, where each variable is set between two thresholds, $Threshold1 < D < Threshold2$. Threshold1 and threshold 2 will be specified depending on the size of the cell.
- Generate the second obfuscation area (R2), which is represented by:

$$(X\sim, Y\sim, Width1 + D1 + D3, Height1 + D2 + D4)$$
$$\text{which is } (X\sim, Y\sim, Width2, Height2)$$

Secondly, the fog node divides the response into five parts (i.e. five separate arrays)

- Normally in LBS the response contains three parts, namely Point of Interest (POI) name, POI latitude, and POI longitude.
- Fog node will split the response into five areas according to their locations (latitude and longitude).

The steps of the proposed DOA algorithm to realize the above architecture are depicted below.

To estimate the query response time of the DOA, let us assume the following:

- Normally, the time of connection to a fog node (using a WIFI connection) takes a 1/10th of the time required to connect to an online SP/Cloud. Therefore, if the connection time to SP = T1, then the time to connect to the fog is going to be $T2 = 0.1 * T1$
- The average time for generating the obfuscation areas R1 and R2 is T3 = less than 1ms.
- The average time for dividing the results into five parts is T4 = 1ms for each 100 results. If the size of obfuscation area increases, T4 will increase accordingly.
- The time of searching in the cache of fog in case of miss T5 is close to 0 because the algorithm applies Bloom filter as Hash-function.
- The time of searching in the cache of fog in case of hit = T_Cache.
- The time of searching in the SP = T_SP.

Therefore, the query response time of DOA in the case of miss-hit = T1 + T2 + T3 + T4 + T5 + T_SP = T1 + 0.1 * T1 + 1 + 1 + 0 = 1.1 * T1 + 2ms + T_SP. So, there is small adverse effect in this case.

However, the time of DOA in the case of cache-hit = T2 + T_Cache + T3 + T4 = 0.1T1 + 2ms + T_Cache. Since the time of searching in the cache < time of searching in the SP, so this will enhance the performance of the system.

The DOA algorithm helps to achieve the following advantages:

1. Integrates various approaches and technologies (i.e. TTP, Obfuscation, Cashing, and Mix-zone) to provide a higher level of privacy than the previous methods individually.
2. Eliminates the need to trust the TTP technology by using a small obfuscation area around the user.
3. Reduces the load on the user by dividing the obfuscation process into two phases.
4. Improves the accuracy of returned results with double obfuscation where only suitably returned areas will be used by the object. This benefit is achieved by dividing the result into five parts; the central one contains the user query along with the first obfuscation area. So the moving object will use the result of the main area as well as the area where it is heading. The remaining areas will be ignored.
5. Increases the sensitivity to new types of attacks such as Correlation Analysis Attack and Inversion Attack (as discussed in the next section).
6. Improves the performance of the whole system by using the fog cache and Bloom filter.
7. Responds to static and dynamic queries of the users.

V. A DOMAIN OF APPLICATION OF THE DOA: CONNECTED VEHICLE SYSTEMS

A. CONNECTED VEHICLES

A MANET is a collection of two or more nodes that are equipped with wireless communication and networking capabilities. These nodes may include laptops, computers, and PDAs, each of which can have a limited transmission range. Nodes can communicate directly if they are located within the transmission range of each other. This type of network is infrastructure-less, self-organizing, adaptive, and does not require any centralized administration. Connected vehicles (i.e. CVs) are a special class of MANET but with some distinctive characteristics, such as high mobility and varying

Algorithm 1 Double Obfuscation Approach (DOA) Algorithm

```
~X: First Obfuscation Area      ~~X: Second Obfuscation      Area OID: Object Identify
FID: Fog Identity               Q: Query of Object SP:       Sever Provider
//Object Function
Input: Level 1 // Integer value in Meters for Obfuscation
Output: Arrays of Results
Start: – e.g. User exists in cell i
While sending a location to LBS is required do
    X ← Current Object's Location
    ~X = Object-A.Create_First_Obfuscation(X, Level1)
    Results = SendToFog(OID,Q, ~X) // Fog of Cell i
    If Object Moves Left Then
            End_Result = Results[0] U Results[1]
      else If Object Moves Up Then
            End_Result = Results[0] U Results[2]
      else If Object Moves Right Then
            End_Result = Results[0] U Results[3]
      else If Object Moves Down Then
            End_Result = Results[0] U Results[4]
      else If Object Moves Up && Left Then
            End_Result = Results[0] U Results[1] U Results[2]
      else If Object Moves Up && Right Then
            End_Result = Results[0] U Results[2] U Results[3]
      else If Object Moves Down && Left Then
            End_Result = Results[0] U Results[1] U Results[4]
      else If Object Moves Down && Right Then
            End_Result = Results[0] U Results[3] U Results[4]
      else
            End_Result = Results[0]
End while
End Function
//Fog Node Function
SendToFog (OID, Q, ~X)
      ~~X = Fog.Create_Second_Obfuscation(~X, Level2)
      Result = SendToSP(FID,Q2, SP))
      Results = Divide_Result(5, Result, ~X, ~~X)
      Return Results
      End
```

Note: The fog node will check its cache for the result of the query before contacting the SP (i.e. caching technology), and the moving object will always use a different ID/Nickname as it moves into a new cell (i.e. mix-zone).

network traffic patterns. CVs are highly dynamic networks where the topology of the network is constantly changing since the vehicles typically move with different speeds [45].

If the communication range between two vehicles is 150m and their speeds are between 60 - 70 mph (25 m/sec), then the presence of the link between the two vehicles would last approximately 12 seconds at most. Vehicles are generally assumed to be equipped with numerous sensors, On-Board-Unit (OBU) and wireless interfaces. Such sensors included global positioning systems (GPS) to provide their precise position as information to routing protocols and speedometers to measure the speeds and directions of the vehicles. According to [46], a vehicle is intelligent if it is equipped with processing, location positioning, and recording capabilities and can operate wireless security protocols.

In addition to the presence of mobile entities (e.g., vehicles), a CVs network also includes road stationary units (i.e., RSUs). For example, traffic lights, road signs, and traffic management systems can be used to provide additional services such as early warning or changing lane notifications. RSUs can be connected to service providers via a cellular network or gateway nodes that provide internet connectivity to the vehicles. CVs provide two types of communication between nodes, namely vehicle to vehicle (V2V) and vehicle to RSU (V2U). Networks in CVs are decentralized and have no fixed infrastructure, so ad hoc nodes depend on themselves for implementing any needed network functionality. As a result, vehicles need to hide driver identities during communication to ensure their privacy [47].

B. FOG COMPUTING

Fog computing is an approach that extends the paradigm of cloud computing to the Internet of Things (IoT) by placing higher-power nodes between end-network devices and the cloud. The concept of fog computing was originally developed by Bonomi as a virtualized platform that can provide key services (i.e., storage, computing, and networking). Fog computing has a similar set of services as cloud computing but includes additional advantages such as close proximity to consumers, dense geographic coverage, and mobility support. The aim of fog computing is to form a layer that provides a real-time and low-latency connection between the edge of network (e.g., traffic lights) and the cloud (e.g., data storage center). Fog computing can support both vertical (i.e., IoT devices to cloud computing) and horizontal (i.e., fog nodes to fog nodes) services [41].

The fundamental element of fog computing is the fog node, where a node can be deployed at different levels and locations between the cloud and edge-network devices to reduce latency, improve the quality of service, and allow real time data analysis. Fog node deployments are sensitive to data processing time. If the generated data is time-sensitive, then the fog node can be implemented close to the edge devices that generate the data (e.g., surveillance cameras). However, if the data is less time-sensitive, then it can be sent to the cloud for historical analysis, such as big data analytics, and long-term storage. As such, IoT data can be directly sent to nearby fog nodes in order to obtain immediate services and then be sent to the cloud servers for future processing [22], [41].

C. ROADSIDE UNITS

RSUs are stationary devices that can be installed inside a roadside electronic cabinet or roadside poles and are assumed to be equipped with storage, processors, and networking capabilities that enable communication to vehicles via the Dedicated Short Range Communication (DSRC) protocol. RSUs units facilitate communication between vehicles and SPs and other devices by transferring data over DSRC in accordance with the industry standards [47].

In figure 5, the smart vehicle hides its real location from the fog, which also adds more obfuscation to hide the location, direction and ID of this moving object from the SP.

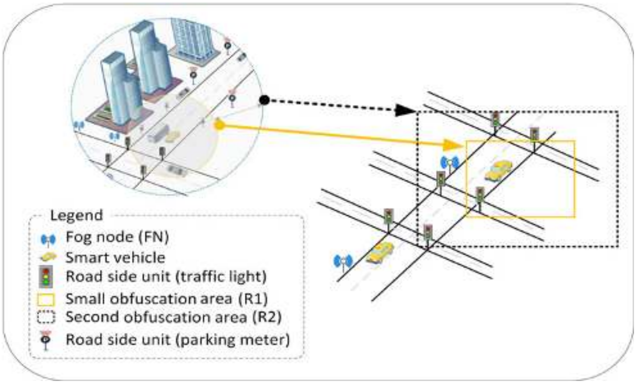

Legend
Fog node (FN)
Smart vehicle
Road side unit (traffic light)
Small obfuscation area (R1)
Second obfuscation area (R2)
Road side unit (parking meter)

FIGURE 5. A possible application of the DOA in the connected vehicle systems.

VI. EVALUATION METRICS AND SIMULATION RESULTS

Our privacy protection approach is based on the principle of obfuscation but in a novel way. To validate the effectiveness of our proposed approach against existing privacy protection techniques, we have conducted two simulation experiments focusing on properties related to obfuscation such as the accuracy of results, overload on the objects, and level of privacy.

The first experiment compares the DOA with the other approaches that merged multi-privacy approaches like the Dummy, Cache, and Peer-Cooperation approach to showcase the differences and demonstrate the advantages offered by our approach. The second experiment, however, focused on identifying the obfuscation issues by shedding light on the Traditional Obfuscation and Enhanced-Obfuscation approaches.

Traditional obfuscation covers the real location of the object within an area or changes it to another dummy location before sending it to the SP. However, Enhanced-Obfuscation is adaptive in nature for it uses correlation to anticipate the new location of the object on the detected path based on the previous queries. Before delving into the details of the simulation experiments, we identified a set of evaluation metrics that were used to guide the comparison of the DOA with other privacy approaches.

A. THE COMPARISON METRICS

The literature proposes numerous criteria to evaluate the quality of privacy protection approaches. Below we focus on three types of metrics to judge the effectiveness of the DOA, namely privacy, performance and logical metrics [19], [20], [25], [28].

1) PRIVACY METRICS

Four quantitative metrics that measure the quality of privacy techniques were chosen for evaluation in the simulations as described below [39], [40], [43], and [44].

1. K-Anonymity: is the percentage of real queries collected by an attacker, compared to the dummy ones. In the DOA, a user does not send her queries to the SP directly; instead, she sends them to the nearest fog node (i.e. TP), which forwards these queries on her behalf. Therefore, all queries that are received by the SP can be considered as dummies. Therefore, the value of this metric will be maximum. K-Anonymity is defined as follows:

$$K\text{-}Anonymity = 1/(1+K) \tag{1}$$

Where K is the number of dummy queries

2. Entropy (denoted E): is the rate of the amount of right information that an attacker can deduce from the collected queries about a specific user. Many other privacy metrics are calculated using Entropy. It is presented by the probability of the degree of certainty. Entropy is defined by the following formula:

$$E = -\sum\nolimits_{i=0}^{n} Pi^{*} Log_2(Pi) \tag{2}$$

where n is the number of queries sent
Pi is the probability that query 'i' belongs to a specific user

In the DOA, the SP cannot determine the identity of the query's owner since all queries come from a third party. Even if the attacker attempts to collect similar queries, its ability is limited by the cell of the TP. Moreover, using obfuscation prevents the attacker from inferring accurate information about the user's location. So the value of Entropy will also be maximum in the DOA, ensuring the highest level of protection.

3. Estimation Error (denoted EE): presents the rate of error that the attacker can fall into. It is defined as follows:

$$EE = (E)^{*}100\% \tag{3}$$

where E is the Entropy value

Privacy techniques aim to maximize this value to enhance the protection level and prevent attackers from detecting the real location of the user. Figure 6 explains the concept of this metric in the obfuscation domain. To increase this value, more noise has to be added (i.e. a larger obfuscation area). However, this will create a negative effect on the accuracy of the results.

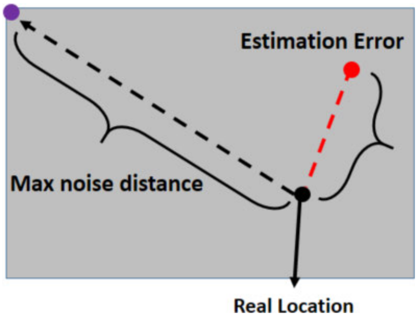


FIGURE 6. The estimation error (EE).

4. Ubiquity (denoted U): refers to the probability of existing in each cell of the area. If user queries originate from different locations in the area, more ubiquity is achieved leading to increased privacy for the user. Ubiquity is defined using the following formula:

$$U = 2^{E} \tag{4}$$

where E is the entropy value.

2) PERFORMANCE METRICS

Similarly, to measure the performance of the DOA, the literature suggests assessing several metrics as listed below [39], [40], [43], and [44].

1. The number of queries (N): this metric refers to the number of queries that are sent to the SP in each sending operation by the user. The DOA uses one query at a time instead of a set of queries as in the Dummy approach, so the number of queries will be kept to a minimum to achieve the best performance.
2. Size of data that are sent in each query or returned in each response: the DOA uses areas instead of a specific location by increasing the radius of each query. This small value (e.g. 2 bytes) will not affect the performance at all.
3. Overhead on the object/user: this metric refers to the size of the obfuscation area or the number of queries generated each time. The DOA divides the obfuscation area into two regions where the user opts for the smaller area to reduce the load.
4. Overhead on the TP: this metric refers to the number of objects and the size of the area. The DOA uses the fog nodes to distribute the density. Each fog node manages a cell that is suitable to its available resources. Moreover, the different sizes of cells address the issue of different density of objects in addition to homogeneity attacks.
5. Overhead on the SP: this metric refers to the amount of information required and the number of queries received. There is no additional overhead compared to the classical obfuscation method. However, the DOA uses caching to reduce the number of connections to the SP, which will reduce the overhead.
6. Processing of the result: most of the privacy-preserving techniques perform some processing before the results are returned to the requester. The obfuscation also needs to do that; however, the fog node divides the area of results into five regions, so the requester (i.e. user) can select one of them according to his current location; this strategy reduces the overhead by more than 50%.
7. Usage of encryption techniques: the DOA does not use any encryption technique for it may adversely influence the performance of the system. If there is any need to use encryption, it will be between user and fog nodes only to reduce its impact on the performance.
8. Average response time: this metric is related to all previous metrics and is usually calculated after testing the system in a real environment or through simulation experiments using representative datasets and comparing the results to other techniques.
9. Cache-hit ratio: some techniques use caching to enhance system performance and privacy; however, the management of cache remains an open challenge. The DOA solves this dilemma by using the cache offered by each fog node. This cache will store only the queries of the fog's cell to increase the rate of hit. Moreover, the DOA applies Bloom filter (i.e. hash function) to eliminate the adverse effect of miss-hit cases in the cache.

3) LOGICAL METRICS

Additional logical criteria were included in the evaluation of the proposed DOA. The integration of several methods to form the DOA enables coping with a wide range of attacks as listed below [20], [26], [41], and [42].

1. Correlation Analysis Attack: in this type of attack the attacker collects the history of locations visited or queries sent by the user (i.e. History Attack) and then analyses it using correlation or regression; for example, the attacker can predict the new location of the user if he has extensive knowledge of the map. This attack poses a real threat to the traditional confusion methods that use fake locations instead of actual locations. This is evaded in the DOA because the attacker will be unable to collect any historical data from a particular user since the user communicates to the service provider indirectly through a third party. Moreover, the user uses two phases of obfuscation in the area, which is harder for the attacker to breach than the traditional obfuscation [19].
2. Tracking Subsequent Obfuscation Areas: this is similar to the previous attack but is more sophisticated. It tracks the path of the user's obfuscation areas instead of the actual query. Also, the attacker does not have any historical records about the user because he does not contact him directly. It also uses the cache, which reduces communication with the service provider and generates gaps that prevent the formation of any path [20].
3. Inversion Attack: in this case the attacker has information about the technology or algorithm used by the user to protect his privacy. This attack can penetrate traditional obfuscation techniques; however, using a third party prevents the attacker or malicious service provider from executing their attacks since the true identity of the authors of the query is concealed [25].
4. Malicious Fog Node or TP Attack: this attack is common in TTP, but the DOA addresses it in two ways; the first way allows the user to apply initial obfuscation before sending the query to a fog node while the second way uses a different alias in each new cell. These two ways prevent the malicious node from revealing the identity of a user or disclosing his exact location during his existence within its cell 41].
5. Kind of Attackers: the DOA does not focus only on a specific type of attackers; instead it takes into consideration all potential attackers including the SP, TP, and outside attackers [26].
6. Type of query that can be supported: the DOA can deal with static queries where the user sends a single query for the POI as well as dynamic queries where the user sends regular queries to the SP during his movement [19], [20].
7. The need for trust: the DOA does not need to trust any third party including the SP, fog node, or other peers within the same cell [25].
8. The accuracy of results: the DOA uses obfuscation to protect user privacy. Normally, the obfuscation technique affects the accuracy of the results. However, the DOA divides the result into five parts before returning them to the user who has to select the suitable parts based on his current location and movement. This will address the issue of accuracy and reduce the overhead on the user [20], [26], and [42].

B. THE FIRST SIMULATION EXPERIMENT

This research presents the simulation results through comparative figures of the aforementioned privacy and performance metrics with a particular focus on the DOA, Enhanced-Cache [38], P2PCache [39], and DCA approach [40]. Microsoft Visual Studio 2015, SQL-Server 2012, and Microsoft Excel 2016 were used to simulate the experiment. This first simulation experiment satisfied the following conditions and requirements:

- The overall test area is divided into 100 * 100 cells
- A fog node exists within each cell in the area. The fog node is considered as a TP with its dedicated cache.
- The size of the Cache is assumed to be 100Kbytes
- The size of each query is assumed to be less than 1Kbytes
- 10000 mobile users are randomly distributed in the cells of the area
- 100 Point of Interests (POIs) are assumed to represent various types of possible queries
- 3G/4G WI-FI connection is available within each cell (i.e. mimicking a smart city environment that has a reasonable IT infrastructure).

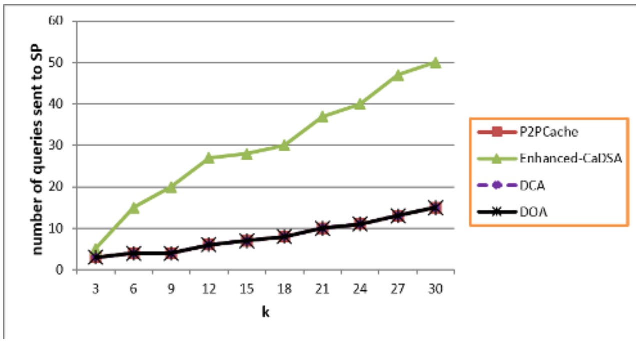


FIGURE 7. Comparison of the number of queries sent to the service provider by the P2PCache, Enhanced-CaDSA, DCA, and DOA, K represents the actual queries of the user.

Figure 7 shows the number of queries sent to the SP compared to the actual number of requested queries (i.e. denoted as K). In the DOA, DCA, and P2PCache the number of queries is considerably less than the number that was sent to the SP due to the use of the cache of the fog nodes. In the worst case scenario, the number of queries will be the same as the number of requested queries because there is no need to use dummies in the proposed technology [38]–[40].

With respect to the cache hit-rate results, the DOA and DCA and P2PCache approach achieved the highest rates since these techniques have a dedicated cache in each cell to store the real queries as opposed to the Enhanced-CaDSA approach which uses dummy queries. Overall, the DOA outperforms the other methods with respect to privacy since it does not need to trust a third party (i.e. a fog node that is responsible for managing the cache). In contrast, the user has to trust the access point manager or other peers in the DCA and P2PCache approaches.

It is known that if the number of queries sent to the SP is reduced, the system privacy and performance are enhanced consequently. The less data that are collected by the SP, the better for the user privacy. Dummy techniques use this strategy by increasing the number of false queries in the SP to reduce the percentage of correct information collected about a specific user. However, user resources are drained with possible negative effects on the response speed. Cache can be used as a technique to reduce the number of queries sent to the SP. Moreover, searching in the cache is faster than the SP and connecting to the fog (i.e. edge computing) is faster than the SP (residing at the cloud).

Our results confirm the previous claims. The DOA, DCA, P2PCache approaches performed better than the Enhance-caDSA, which uses dummy queries negatively impacting the cache hit-ratio. The DOA, DCA, and P2PCache approaches send one real query only resulting in high cache hit-ratio. In our experiment, we have assumed that the cache size of all approaches (DOA, DCA, P2PCache, and Enhance-caDSA) is equal. However, the DOA emerged as a better method with respect to managing the cache by exploiting the fog in available within each cell. This is contrary to the P2PCache which uses the cache of the phone devices of the user and to the DCA which uses general access points. Finally, we repeated each K queries ten times, then calculated the average number of queries sent for each selected K queries.

The results show that 12 out of 30 (40%) requested queries were answered by cache without the need to forward them to the SP, thus improving the user privacy and system performance.

Figure 8 shows the average response time to the queries calculated from the moment of sending the queries until receiving the results by the user. Again, we repeated each set of queries (N) ten times, and then calculated the average time to enhance the accuracy. We note that the DOA outperforms the Enhanced-CaDSA technology because the Enhanced-CaDSA relies on sending a lot of dummies along with the real query which causes additional time. The DOA relies on the fog node to reduce the needed time to create the obfuscation zone and process the results.

Notably, the DOA responded faster than the DCA and P2PCache in the case of many queries. That is due to the fact that dealing with the fog node, which acts as the TP, is faster than dealing with the peers in the DCA and P2PCache.

The results depicted in Figure 8 shows that cooperating with fog nodes (i.e. using the DOA) is more performant than generating many dummies (the case of Enhance-caDSA) or cooperating with other moving peers (the case of DCA or P2PCache). Actually, generating an obfuscation area is created using a simple function, whilst processing the results of the fog nodes after obfuscation can impose a significant load on the user. For this reason, the DOA splits the results provided by fog nodes into five parts without preaching user privacy.

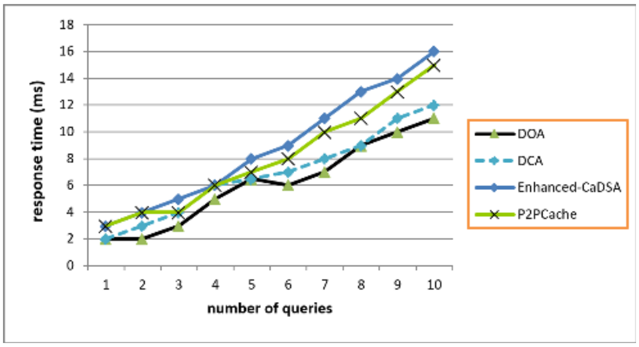


FIGURE 8. A comparison of the response rime (in milliseconds) to the number of queries sent by the P2PCache, Enhanced-CaDSA, DCA, and DOA.

The results show the average response time for handling 10 queries is approximately 11 milliseconds, which is faster than the other approaches since creating an obfuscation area is faster than generating many smart dummies or searching for one or more cooperator peers. Moreover, about 30% of the queries are served by the fog cache without the need to connect to the SP.

Figure 9 shows the amount of additional results that will be received by the user due to the obfuscation area (i.e. the dotted circle). The user has to filter all results which causes additional load on her resources. To overcome this issue, the DOA uses the fog nodes to filter the results and split the results into five parts.

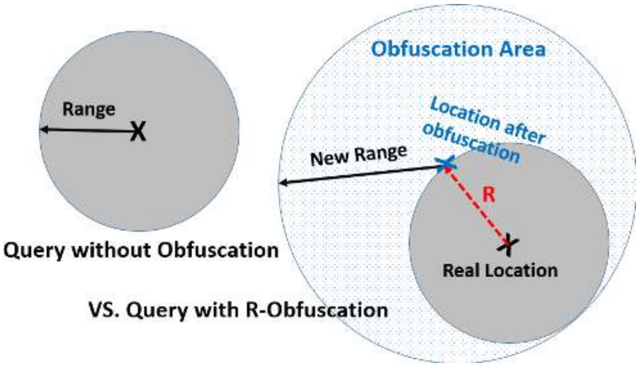

Range
Query without Obfuscation
Obfuscation Area
Location after obfuscation
New Range
R
Real Location
VS. Query with R-Obfuscation

FIGURE 9. A comparison of queries with and without obfuscation.

Figure 10 shows the level of privacy using the entropy metric. This metric signifies the amount of data collected by the SP, where the SP ensures that these data are related to a specific user. Normally, the entropy value should decrease with each new query sent by the user since the SP will collect more information which increases the certainty about the user identity. Dummy techniques mix the query with many false

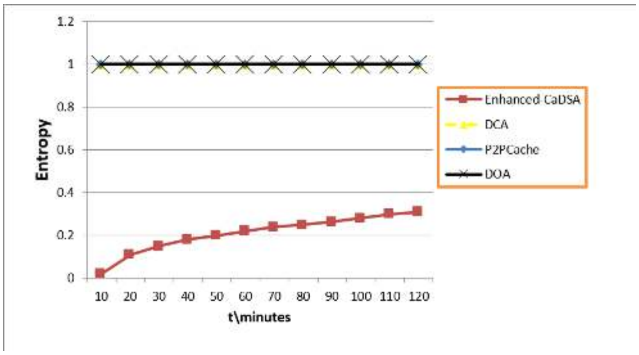


FIGURE 10. A comparison of entropy values by the Enhanced-CaDSA, P2PCache, DCA, and DOA, t represents time in minutes.

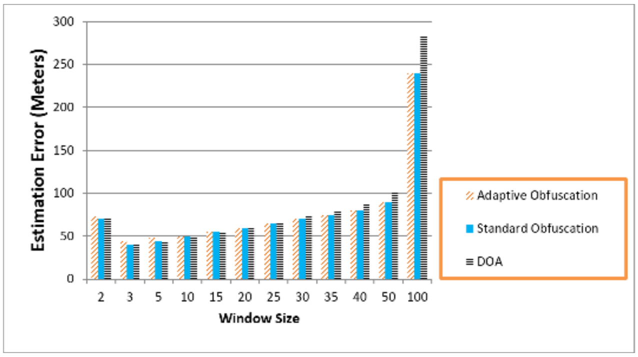


FIGURE 11. A comparison of the estimation error (in meters) by the adaptive obfuscation, standard obfuscation and DOA.

queries to mislead the SP. The DOA achieved the best entropy. This result is similar to the DCA and P2PCache approaches if we measure this metric in the SP. This can be justified by the fact that these approaches do not communicate directly with the SP. This means that the SP is unable to create true historical data for the users using these approaches. However, the DOA is considered superior because there is no need to trust the TP as opposed to the P2Pcache or DCA approaches [38]–[40].

Using the DOA, DCA, and P2PCache approaches, the user does not connect directly to the SP, resulting in maximum protection of the user privacy. The SP will not be able to link the received queries from the fog node to any user. The entropy will be maximum (i.e. E = 1) in all of these three approaches. However, in the DCA and P2PCache approaches, a peer ''A'' will send her query of another peer ''B'', so the query of B will act as a dummy for A and vice versa. The location of peer A is usually close to peer B in peer cooperation approach. So DCA and P2PCache may suffer from a breach of the peer's location. This problem is adequately addressed by the DOA using the obfuscation concept.

C. THE SECOND SIMULATION EXPERIMENT

In the second simulation experiment, a Geo-life dataset set was randomly selected. This set contained approximately more than 17000 GPS trajectories of 182 unique users over the past 3 years. The experiment used some trajectories of real users to test the estimation error percentage across the DOA, standard (i.e. classic) obfuscation [42], and adaptive obfuscation approach [44]. The estimation relies on the correlation factor of the previous locations. The estimate error refers to the amount of errors made when the attacker attempts to detect the users' real location after applying the obfuscation.

Figure 11 shows the tradeoff between the size of obfuscation window (i.e. the area of obfuscation) and system performance. Larger obfuscation windows achieve more privacy for the user's location, but at the same time it causes more overload on the user's resources. In the worst-case scenario, the DOA will give exactly equal results to the traditional obfuscation technique with respect to the amount of error that the attacker can fall into. However, the adaptive obfuscation technique is anticipated to create a more erroneous guess rate. As stated earlier, the user of the DOA does not deal with the SP since a fog node will hide the user's identity and send it on his behalf.

The DOA creates two areas obfuscation by two different parties (i.e. the user and fog) and then the fog will filter the results so there is no harm in generating large areas. In fact, the best results are achieved by the DOA when using large windows. Small obfuscation windows do not take advantage of the two phases of obfuscating or the division of results.

Overall, the DOA achieves better privacy compared to the other techniques of obfuscation (e.g. traditional and adaptive), as depicted in Figure 11. The error of estimation by the attacker will increase due to the two random phases of obfuscation. For a window size of 100m, the estimation error was about 180m, where the window size refers to the obfuscation distance between real position of the user and the obfuscated position.

Moreover, the process of obfuscation itself is divided into two phases separating the user and the fog (i.e. the third party). Finally, using the cache reduces the number of communications required with the SP. The proposed DOA outperformed all previous methods of obfuscation including the adaptive obfuscation, as demonstrated by the bigger window sizes (i.e. obfuscation areas). This is because the DOA uses two phases to create the obfuscation areas resulting in higher randomization rates and estimation errors. Moreover, the DOA presents a solution to the accuracy problem of the result, which is not addressed by the previous obfuscation techniques. The DOA also eliminates the issue of overloading the user by using the fog node to create the obfuscation area and handling the response to and from the users.

D. THE THIRD SIMULATION EXPERIMENT

We have conducted a third experiment to explore the effects of the caching technology offered by fog computing on the privacy and performance of the DOA.

In fact, several studies investigated the use of caching technology to improve the system performance. In particular,

TABLE 2. A comparison of the proposed DOA against major privacy techniques.

Privacy Technique	No need to trust the SP	No need to trust the TP	Good Performance	Does not overload user	Accurate Results	Privacy of ID	Privacy of Query	Privacy of Location
Enhanced-CaDSA	✓	✓			✓		✓	✓
P2PCache	✓		✓	✓	✓	✓	✓	
DCA	✓		✓	✓	✓	✓	✓	
Standard Obfuscation	✓	✓						✓
Adaptive Obfuscation	✓	✓						✓
Double Obfuscation Approach (DOA)	✓	✓	✓	✓	✓	✓	✓	✓

it has been used mainly to answer future user queries without connecting to the SP, which improves the privacy level of the users. Typically, the cache is used in the central TTP, but recent research approaches, e.g. CAST [34] and P2PCache [39], exploited the resources of users' devices to enable a seamless integration between the cache and cooperating peers. Nonetheless, this concept suffered from several issues including disconnection, overhead on the user device, and need for trust between the peers. Alternative methods, such as Enhance-CaDSA [38] and PPCS [48], addressed these issues by dividing the area to many cells where each cell has one dedicated cache; however, there is no clear management strategy to administer these dedicated caches in the system.

The DOA uses the cache that comes ready within the fog nodes, where each fog will play the role of a TTP in its cell and will manage and protect its cache. Since the DOA does not use dummy queries, all stored queries will be real which will in turn enhance the cache-hit ratio. There is a high probability that other users will request the same queries if they were in same area [50]. To depict the cache advantages on our system, we conducted this simulation that measures the privacy and performance level of the system based on different values of the cache-hit ratio. Firstly, we will outline some basic assumptions.

- The time for sending the query to the SP without protection will be T1 = A. For example, A = 20ms (milliseconds) using a 4G connection, as a Ping test to the SP.
- The time for sending the query to the fog node in the worst-case scenario will be T2 = B. For example, B = 4ms (milliseconds) using a WIFI connection.
- The time taken by the DOA to satisfy the query when the result is fetched from the cache of fog node would be T3 = B.
- The time taken by the DOA to satisfy the query when the result is not present in the cache of the fog node would be T4 = T1 + T2
- If H represents the cache hit-ratio, then the total time for processing N queries will be equal to N * T3 * H + N * T2 * (1 − H), where N * (1 − H) represents the number of queries that will be sent to the SP.

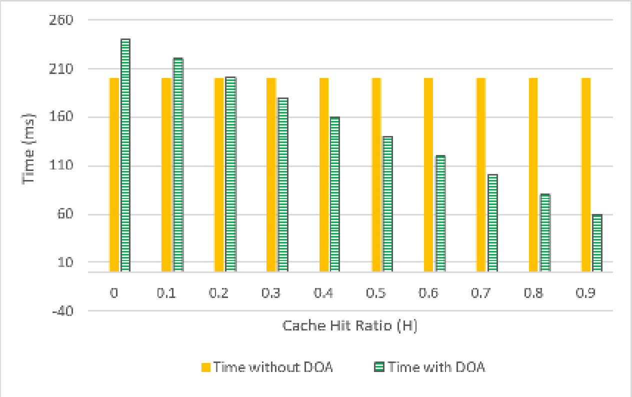


FIGURE 12. Effects of cache hit-ratio on the response time (i.e. system performance) with and without the DOA.

Figure 12 shows the time (in milliseconds) it takes to process user queries with and without using the DOA. Overall, it can be seen that as the cache hit ratio increases the response time decreases continually. It is worthwhile to note that when the cache hit ratio =< 0.2, a delay is introduced to the DOA as a result of connection to the SP to fetch the responses; however, as soon as the hit cache ratio surpasses 0.2 (H > 0.2), the DOA reduces the response time of the system significantly since the responses would have been cached by the fog nodes already.

Figure 13 shows the relationship between the cache hit ratio and the number of queries sent to the service provider (SP) with and without using the DOA. Overall, it can be observed that as the cache hit ratio increases the number of queries sent by the DOA decreases significantly. When the DOA is not used, the number of queries communicated to the SP remains the same. It is evident that higher cache hit ratios contribute to less queries being sent to the SP, which works in favor of protecting the user privacy.

E. OVERALL COMPARISON

Finally, Table 2 summarizes the key findings of the simulation experiments with respect to the Enhanced-CaDSA,

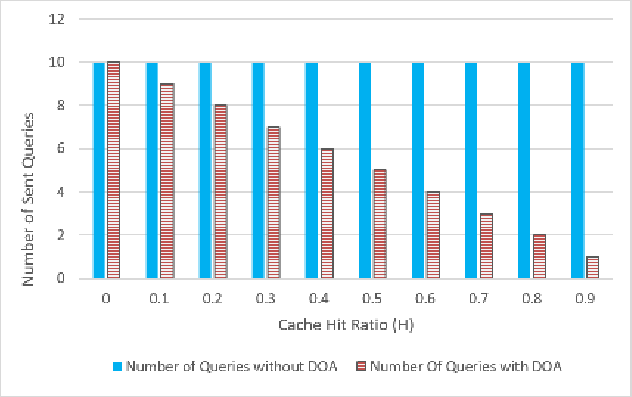


FIGURE 13. **Effects of cache hit-ratio on the number of queries sent to the SP (i.e. user privacy).**

P2PCache, DCA, Standard Obfuscation, Adaptive Obfuscation, and DOA approaches.

VII. CONCLUSION AND FUTURE WORKS

This research proposed a newly developed approach called Double Obfuscation Approach (DOA), which integrates two privacy protection approaches (i.e. Obfuscation and TTP) and enhances their capabilities by exploiting two technologies (i.e. Caching and Mix-Zone). The DOA achieved a greater level of privacy and succeeded in addressing several problems of previous privacy protection approaches such as the overhead, need to trust a third party and accuracy of the returned results.

The DOA combines two interesting concepts; firstly, it divides the obfuscation area into two phases, which are supervised by the user and the TP (i.e. fog node) on behalf of the user. Secondly, when the TP sends a query to the SP and receives the result, the DOA divides this result into five parts to enable the user to deal with them easily and select the appropriate parts according to his current location and direction. Finally, the concept of Mix-Zone is used to prevent the SP and malicious TPs from tracing the path of the user movement. Moreover, using the cache enhances the performance and reduces the number of connections to the SP. The results of our simulations demonstrated the superiority of the DOA with regards to privacy, accuracy and overhead.

However, there are various qualifying limitations of the proposed approach that should be highlighted and explored in the future. Specifically, the DOA has not investigate the possibility of a large number of malicious fog nodes (TPs) cooperating with the malicious SP, although this is unlikely to happen. We anticipate that the use of the Mix-Zone concept in the DOA, where users change their nicknames in every new cell, can solve this issue especially when there is a large number of users within each cell. However, further testing is required to assert this assumption. We plan to carry out more experiments to measure the achieved level of privacy using various quantitative measures, such as Entropy, Estimated Error, F-Measure, . . . etc., [20], [44], [48].

REFERENCES

[1] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Comput. Netw.*, vol. 54, no. 15, pp. 2787–2805, Oct. 2010.

[2] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A survey on Internet of Things: Architecture, enabling technologies, security and privacy, and applications," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1125–1142, Oct. 2017.

[3] A. A. Mutlag, M. K. 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.

[4] M. Bembe, A. Abu-Mahfouz, M. Masonta, and T. Ngqondi, "A survey on low-power wide area networks for IoT applications," *Telecommun. Syst.*, vol. 71, no. 2, pp. 249–274, Jun. 2019.

[5] A. Narmada and P. S. Rao, "RFID integration with wireless sensor networks," *Res. J. Eng. Technol.*, vol. 9, no. 2, pp. 207–213, 2018.

[6] Y. Zhao and N. Patwari, "RFID localization in wireless sensor networks," in *Radio Frequency Identification*, vol. 19. London, U.K.: IntechOpen, 2017.

[7] A. Yassine, S. Singh, M. S. Hossain, and G. Muhammad, "IoT big data analytics for smart homes with fog and cloud computing," *Future Gener. Comput. Syst.*, vol. 91, pp. 563–573, Feb. 2019.

[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.

[9] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "Connected vehicles: Solutions and challenges," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 289–299, Aug. 2014.

[10] A. Talebpour and H. S. Mahmassani, "Influence of connected and autonomous vehicles on traffic flow stability and throughput," *Transp. Res. C, Emerg. Technol.*, vol. 71, pp. 143–163, Oct. 2016.

[11] K. Su, J. Li, and H. Fu, "Smart city and the applications," in *Proc. Int. Conf. Electron., Commun. Control (ICECC)*, Sep. 2011, pp. 1028–1031.

[12] X. Zhu and C. Zhou, "POI inquiries and data update based on LBS," in *Proc. Int. Symp. Inf. Eng. Electron. Commerce*, May 2009, pp. 730–734.

[13] K. Sonklin, Y. Feng, D. Jayalath, and C. Wang, "A new location-based services framework for connected vehicles based on the publish-subscribe communication paradigm," MDPI, Basel, Switzerland, Tech. Rep., 2019.

[14] N. Alhalafi and P. Veeraraghavan, "Privacy and security challenges and solutions in IOT: A review," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 322, Sep. 2019, Art. no. 012013.

[15] J. Veijalainen, D. Kozlov, and Y. Ali, "Security and privacy threats in IoT architectures," in *Proc. 7th Int. Conf. Body Area Netw.*, Feb. 2012, pp. 256–262.

[16] 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.

[17] R. H. Weber, "Internet of Things—New security and privacy challenges," *Comput. Law Secur. Rev.*, vol. 26, no. 1, pp. 23–30, Jan. 2010.

[18] R. Rodrigues, D. Barnard-Wills, P. De Hert, and V. Papakonstantinou, "The future of privacy certification in europe: An exploration of options under article 42 of the GDPR," *Int. Rev. Law, Comput. Technol.*, vol. 30, no. 3, pp. 248–270, Sep. 2016.

[19] R. Gupta and U. P. Rao, "An exploration to location based service and its privacy preserving techniques: A survey," *Wireless Pers. Commun.*, vol. 96, no. 2, pp. 1973–2007, Sep. 2017.

[20] A. A. A. Sen, F. A. Eassa, K. Jambi, and M. Yamin, "Preserving privacy in Internet of Things: A survey," *Int. J. Inf. Technol.*, vol. 10, no. 2, pp. 189–200, 2018.

[21] A. Pravin, P. Jacob, and G. Nagarajan, "A comprehensive survey on edge computing for the IoT," in *Edge Computing and Computational Intelligence Paradigms for the IoT*. Hershey, PA, USA: IGI Global, 2019, pp. 37–45.

[22] Y. Qu, M. R. Nosouhi, L. Cui, and S. Yu, "Privacy preservation in smart cities," in *Smart Cities Cybersecurity and Privacy*. Amsterdam, The Netherlands: Elsevier, 2019, pp. 75–88.

[23] Y. Chen, J. Bao, W.-S. Ku, and J.-L. Huang, "Cache management techniques for privacy preserving location-based services," in *Proc. 9th Int. Conf. Mobile Data Manage. Workshops, MDMW*, Apr. 2008, pp. 88–96.

[24] B. Farahani, F. Firouzi, V. Chang, M. Badaroglu, N. Constant, and K. Mankodiya, "Towards fog-driven IoT eHealth: Promises and challenges of IoT in medicine and healthcare," *Future Gener. Comput. Syst.*, vol. 78, pp. 659–676, Jan. 2018.

[25] 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.

[26] M. Wernke, P. Skvortsov, F. Dürr, and K. Rothermel, "A classification of location privacy attacks and approaches," *Pers. Ubiquitous Comput.*, vol. 18, no. 1, pp. 163–175, Jan. 2014.

[27] S. N. Matheu, A. R. Enciso, A. M. Zarca, D. Garcia-Carrillo, J. L. Hernández-Ramos, J. Bernal Bernabe, and A. F. Skarmeta, "Security architecture for defining and enforcing security profiles in DLT/SDN-based IoT systems," *Sensors*, vol. 20, no. 7, p. 1882, Mar. 2020.

[28] M. S. Alrahhal, M. Khemakhem, and K. Jambi, "A survey on privacy of location-based services: Classification, inference attacks, and challenges," *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 24, pp. 6719–6740, Dec. 2017.

[29] I. B.-P. K. Gudymenko and K. Tietze, "Privacy implications of the Internet of Things," in *Proc. Int. Joint Conf. Ambient Intell.*, Berlin, Germany: Springer, Nov. 2011, pp. 280–286.

[30] H. Kido, Y. Yanagisawa, and T. Satoh, "An anonymous communication technique using dummies for location-based services," in *Proc. Int. Conf. Pervasive Services ICPS*, Jul. 2005, pp. 88–97.

[31] A. ?Solanas, J. Domingo-Ferrer, and A. Martínez-Ballesté, "Location privacy in location-based services: Beyond TTP-based schemes," in *Proc. 1st Int. Workshop Privacy Location-Based Appl. (PILBA)*, Oct. 2008, pp. 12–23.

[32] S. T. Peddinti and N. Saxena, "On the limitations of query obfuscation techniques for location privacy," in *Proc. 13th Int. Conf. Ubiquitous Comput. UbiComp*, 2011, pp. 187–196.

[33] A. Khoshgozaran and C. Shahabi, "Private information retrieval techniques for enabling location privacy in location-based services," in *Privacy in Location-Based Applications*. Berlin, Germany: Springer, 2009, pp. 59–83.

[34] R. Gupta and U. P. Rao, "Achieving location privacy through CAST in location based services," *J. Commun. Netw.*, vol. 19, no. 3, pp. 239–249, 2017.

[35] 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. (GLOBECOM)*, Dec. 2013, pp. 820–825.

[36] S. Zhang, Q. Liu, and G. Wang, "A caching-based privacy-preserving scheme for continuous location-based services," in *Proc. Int. Conf. Secur., Privacy Anonymity Comput., Commun. Storage*, Cham, Switzerland: Springer, Nov. 2016, pp. 73–82.

[37] H. Ngo and J. Kim, "Location privacy via differential private perturbation of cloaking area," in *Proc. IEEE 28th Comput. Secur. Found. Symp.*, Jul. 2015, pp. 63–74.

[38] 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.

[39] 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.

[40] 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.

[41] A. Alrawais, A. Alhothaily, C. Hu, and X. Cheng, "Fog computing for the Internet of Things: Security and privacy issues," *IEEE Internet Comput.*, vol. 21, no. 2, pp. 34–42, Mar. 2017.

[42] C. A. Ardagna, M. Cremonini, E. Damiani, S. D. C. Di Vimercati, and P. Samarati, "Location privacy protection through obfuscation-based techniques," in *Proc. IFIP Annu. Conf. Data Appl. Secur. Privacy*, Berlin, Germany: Springer, Jul. 2007, pp. 47–60.

[43] G. Theodorakopoulos, "The same-origin attack against location privacy," in *Proc. 14th ACM Workshop Privacy Electron. Soc. WPES*, 2015, pp. 49–53.

[44] R. Al-Dhubhani and J. M. Cazalas, "An adaptive geo-indistinguishability mechanism for continuous LBS queries," *Wireless Netw.*, vol. 24, no. 8, pp. 3221–3239, Nov. 2018.

[45] J. Harding, G. Powell, R. Yoon, J. Fikentscher, C. Doyle, D. Sade, M. Lukuc, J. Simons, and J. Wang, *Vehicle-to-vehicle communications: Readiness of V2V technology for application (No. DOT HS 812 014). United States*, Washington, DC, USA: National Highway Traffic Safety Administration, 2014.

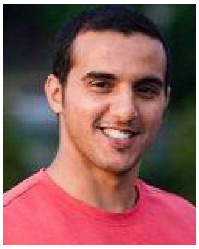
[46] A. Alnasser, H. Sun, and J. Jiang, "Cyber security challenges and solutions for V2X communications: A survey," *Comput. Netw.*, vol. 151, pp. 52–67, Mar. 2019.

[47] J. E. Siegel, D. C. Erb, and S. E. Sarma, "A survey of the connected vehicle landscape–architectures, enabling technologies, applications, and development areas," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 8, pp. 2391–2406, Aug. 2018.

[48] 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.

[49] 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.

[50] R. Shokri, G. Theodorakopoulos, P. Papadimitratos, E. Kazemi, and J.-P. Hubaux, "Hiding in the mobile crowd: LocationPrivacy through collaboration," *IEEE Trans. Dependable Secure Comput.*, vol. 11, no. 3, pp. 266–279, Jun. 2014.



SAMI SAAD ALBOUQ received the B.S., M.S., and Ph.D. degrees in computer science, in 2009, 2011, and 2018, respectively. He is currently an Assistant Professor with the Computer Science College and the Dean of Information Technology Deanship with the Islamic University of Madinah, Almadinah, Saudi Arabia. His research interests include system security engineering, complex information technology systems, and quantum key distribution systems.

ADNAN AHMED ABI SEN received the Ph.D. degree in computer science from the King Abdulaziz University of 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 the Islamic University of Saudi Arabia and a Lecture of Network Security Diploma program. He is also an Experienced and an Established Researcher in privacy and mobile computing. He also has good expertise in systems analyzes and development. He has published about 20 research articles, many of which are indexed in Thomson and Reuters database.

ABDALLAH NAMOUN received the bachelor's degree in computer science and the Ph.D. degree in informatics from the University of Manchester, U.K., in 2004 and 2009, respectively. He is currently an Associate Professor of interactive systems and the Head of the Information Technology Department, Faculty of Computer and Information Systems, Islamic University of Madinah, Saudi Arabia. He has authored more than 50 publications in research areas spanning software engineering, human–computer interaction, and intelligent systems. He investigates user needs and interaction with modern interactive technologies, design of composite software services, and novel methods for testing the usability and acceptance of human interfaces. His recent research interests focus on integrating state of the art artificial intelligence techniques in the design of intelligent user interfaces and the design of Arabic user experience.

NOUR MAHMOUD BAHBOUH received the bachelor's degree from the Faculty of Information Technology Engineering, Al-Baath University, Homs, Syria, in 2010, and the master's degree in web sciences from SVU. She is currently a Lecture with Islamic University, Cipher Security Diploma. Her research interests include privacy, the IoT applications, and education issue.

AHMAD B. ALKHODRE received the B.Eng. degree in computer engineering from the University of Aleppo, Syria, in 1995, and the master's and Ph.D. degrees in computer science from the CITI Laboratory, Communicated Embedded Systems Group, INSA- Lyon, France, in 2000 and 2004, respectively. He is currently a Professor with the Department of Information Technology, Islamic University of Medina. His main interests include sensing networks and security of embedded systems.

ABDULLAH ALSHANQITI received the B.Sc. degree in computer science from Taibah University, Madinah, Saudi Arabia, and the M.Sc. and Ph.D. degrees from the University of Leicester, U.K. In 2012, he joined the Faculty of Computer Science and Information Technology, Islamic University of Madina, as a Lecturer, and was appointed as an Assistant Professor in smart systems and software reverse engineering in 2018. He is currently recognized for his work on software reverse engineering based on dynamic analysis and model/graph transformations using intelligent learning and inference approaches. He is interested in extending his knowledge by welcoming any research cooperation in different cutting-edge disciplines, including quantum machine learning, hybrid AI approaches that focus on solving NLP challenges, and interpretability of deep learning models using graph transformations rules.