IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 3 June 2024, accepted 20 June 2024, date of publication 27 June 2024, date of current version 29 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3420248

RESEARCH ARTICLE

Double Cloak Area Approach for Preserving Privacy and Reliability of Crowdsourcing Data

NOUR MAHMOUD BAHBOUH[1], AHMAD B. ALKHODRE[2], SANDRA SENDRA[3], (Member, IEEE), ADNAN AHMED ABI SEN[2], YAZED ALSAAWY[2], MOHAMED BENAIDA[2], AND HANI ALMOAMARI[2]

[1]Department of Information and Communication Sciences, Granada University, 18071 Granada, Spain
[2]Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah 42351, Saudi Arabia
[3]Instituto de Investigación para la Gestión Integrada de Zonas Costeras, Universitat Politècnica de València, 46730 Valencia, Spain

Corresponding authors: Ahmad B. Alkhodre (aalkhodre@iu.edu.sa) and Adnan Ahmed Abi Sen (Adnanmnm@hotmail.com)

This work was supported by the Research Deanship of Islamic University of Madina under Grant 966.

ABSTRACT Crowdsourcing has emerged as a pivotal data source for diverse smart city applications, ranging from health and traffic to security and safety. However, the integration of users' location data in crowdsourced information poses a significant privacy challenge. Current privacy protection approaches of location-based services have become inadequate to face the evolving attackers' techniques and tools. Moreover, these protection methods ignored the issue of preserving the accuracy and reliability of data. This paper introduces a novel approach, termed Double Cloak Area (DCL-Ar), designed to effectively safeguard users' location privacy and ensure the reliability of data based on crowdsourcing. DCL-Ar differentiates by offering dual-layer protection for identity. The first layer involves users creating an initial cloak zone, while the second layer utilizes fog nodes to establish an extended cloak zone. Furthermore, the proposed method introduces three distinct scenarios for managing collaboration among fog nodes to select the optimal anonymizer and address the limitations of existing protection methods which are related to saving the reliability and the accuracy of data. DCL-Ar maintains maximum entropy, achieving complete uncertainty about user locations, thereby ensuring a high level of privacy protection. Through simulation and comparative analysis, the efficacy of the proposed approach is demonstrated where it provides a superior privacy level without significant performance. Experimental results demonstrate that DCL-Ar outperforms traditional methods, improving cache hit ratios and response times while reducing server query loads. Specifically, our approach reduces the number of queries sent to the service provider (SP) by up to 50% compared to existing methods and maintains a high cache hit ratio of nearly 100% over time. It further impacts on the traditional cloak-area and other protection approaches.

INDEX TERMS Anonymizer, crowdsource, fog, privacy, smart city.

I. INTRODUCTION

The Internet of Things (IoT) has changed many concepts of our lives and made our cities smarter [1]. In achieving this, the IoT relied mainly on sensing data from every-where through billions of wireless network sensors (WSNs) and Radio Frequency Identification (RFID) [2]. The WSNs provide information about the surrounding environment, such as temperature, pressure, noise, pollution level, etc. [3]. RFID tags provide unique object identifiers that enable systems and applications to interact and track them [4].

In general, the huge amounts of data generated by the IoT (through the layer of sensors, tags, and smart devices) is the main axis in creating systems and services that are more advanced and adapted by the users [5]. Because of the weakness of the resources for the IoT, it is necessary to rely on analyzing the data within the cloud [6]. This step helps to understand the behavior of users and provides services that are more commensurate with their requirements, detects services' defects and does remedy them, or reveals

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

new knowledge and supports more accurate and correct decisions [7].

The cloud is not able to meet the requirements of all types of smart systems and services on its own, especially those that are sensitive to delays, such as medical applications, traffic congestion, disaster handling, and others [8], [9], [10]. Therefore, new computing paradigms have emerged, such as edge computing or what is known as fog computing [11], where smart cities provide a network of fog nodes spread densely to cover most areas of the city in the form of adjacent cells. The fog node presence near IoT tools and users makes it able to conduct rapid processing in real-time and provides alerts for immediate responses in emergencies. Thus, cloud computing and fog computing integration has contributed to supporting more forms of smart applications [12]. Figure 1 shows the main layers in the IoT and the stages and processes in each layer [13].

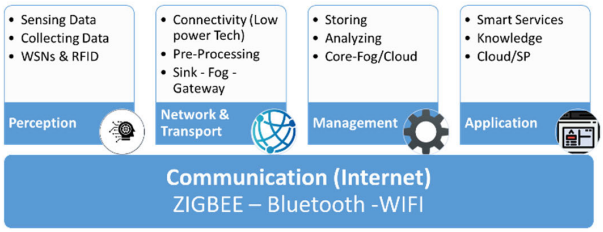

FIGURE 1. Main layers and phases of IoT.

Recently, new models such as the Crowdsourcing have emerged to provide better data than the IoT in many cases [14]. Crowdsourcing integrates human perception, experiences, and evaluation with visual and auditory recognition on one hand, in addition to the sensors and technologies in smart devices on the other hand [15]. Moreover, mobile phones have become more resourceful, so the devices' resources can be used to carry out primary data processing and reduce the load on the higher computing layers [16]. Despite the importance of new models of data sources, such as crowdsourcing, in supporting smarter and more effective services, they face challenges related to the reliability of the transmitted data, the containment of malicious and misleading data, and the challenge of the security and privacy, especially with the adoption of crowdsourcing models on the location of users as one of the main parts of the data sent [17], [18].

In recent years, interest in protecting the privacy and security of users has increased dramatically. Most developed countries have developed privacy laws, such as the European law General Data Protection Regulation (GDPR) and the US law [19]. These laws focused specifically on managing the relationship of service providers (SPs) with their users' data. Unfortunately, most of the data coming from the IoT or Crowdsourcing models can be analyzed to reveal a lot of sensitive data about its users. It is even more dangerous when this data contains the user's location, such as where he is at certain times or what he usually visits. It is necessary that the data contains the user's location for Location Based Services (LBS) and smart systems [20].

LBS represents a sizable leap in the level and a form of new services and applications, especially with the large spread of smart devices and mobile phones, in addition to the development in communication technologies [21]. But revealing users' location compounds the risk for both privacy and security. Figure 2 shows a comparison between security and privacy concerns [22].

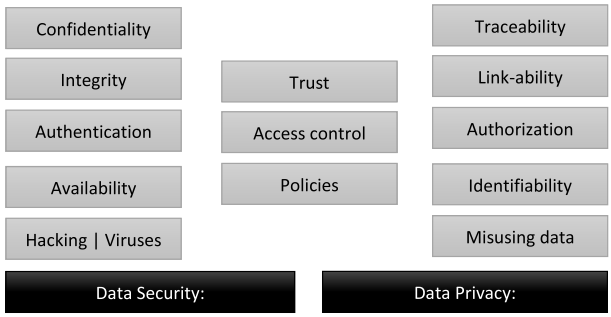

FIGURE 2. Security vs. privacy.

Despite the interdependence between the concepts of privacy and security, there is a big difference in interest in each of them, or what is known as the triple interest. Firstly, data security is concerned with confidentiality and not exposing data to any unauthorized one by encrypting it. Secondly, data security is concerned with ensuring the integrity and reliability of data and not modifying it during transmission or storage over the network. Thirdly, data security is concerned with ensuring data availability permanently, continuity of the service, and not being subjected to an attack that causes it to stop [23].

On the other hand, privacy is about users. Firstly, privacy seeks to hide the owner's identity of transmitted or stored data over the network. Moreover, privacy limits the ability of SPs to link data to users and create a profile for each user, thus revealing additional data about him. Finally, privacy is keen to prevent users from being tracked, such as knowing their locations over time [24].

The attacks that LBS can suffer from, can be categorized into two types of attacks; the anonymous attack, and the active attack. As for the anonymous attack, the attacker seeks to steal users' data, analyzes it, and reveals a lot of sensitive data about them without the users' knowledge. While in the active attack, the attacker seeks to block the service or modify the data. The first attack (anonymous) may be more dangerous than the second because the user is unaware of the problem. The first attack is also more privacy-related, especially in forms like Crowdsourcing. A single piece of information sent by the user may not be confidential, but collecting a lot of information and linking it to the user and his location over time will cause a real threat to reveal much of the user's privacy, such as information about user's habits, behavior, religion, work, social status, and others [25].

Therefore, in this proposed work, we focus on protecting privacy within Crowdsourcing applications from the SP, knowing that the malicious SP is more dangerous than the external attacker, as it has access to all the stored data. Moreover, this work provides a mechanism to improve the reliability of the data and proposes an idea to improve the level of security as well.

Data privacy and security constitute the biggest challenge for users to cooperate with modern services and systems, especially those that use user data sent as a basis, such as Crowdsourcing and LBS models.

Briefly, the contributions to this research work are as follows:

- Proposed an improved approach to ensure users' privacy in Crowdsourcing-based services.
- Employed a collaboration between fog nodes in smart cities to improve privacy and ensure data reliability in crowdsourcing models.
- Proposed three scenarios to manage the collaboration process at the peers' level and the fog nodes level.
- Presented real cases of applying the proposed approach in different smart applications and services.
- Implemented a simulation to prove the superiority of the proposed approach over previous privacy methods.

This paper proposes a new method called Double Cloak Area (DCL-Ar) to protect user privacy in smart city applications that rely on crowdsourced data. This method combines user-defined privacy zones with additional layers created by devices called fog nodes. This double layer of protection keeps user data private while still ensuring the data is accurate and usable. Tests show that DCL-Ar is better at protecting privacy than other methods, making it a valuable tool for balancing privacy and security in smart cities without affecting how well services work.

The second part of the research is a reference study. The third section details the proposed approach. The fourth section presents several cases of applying the proposed approach in different applications. The fifth section shows the results of the simulation and a comparison between the proposed approach and previous methods. Finally, a conclusion that highlights some points for the future.

II. RELATED WORK

This section conducts an extensive examination of pertinent studies focusing on privacy preservation, with a specific emphasis on LBS, particularly those employing Crowdsourcing models for data collection. Additionally, the section conducts a comparative analysis of preceding methodologies, delineating their respective advantages and disadvantages. Furthermore, the scrutiny delineates the specific facets of user data or queries safeguarded by each methodology. Moreover, a succinct overview is presented in Table 1 for expeditious reference.

A. ANONYMITY

This is the simplest way to protect privacy. It depends on concealing the user's identity by replacing it with a pseudonym or a fictitious or specific code so that the attacker cannot link the collected data to the user's file. That means that the obtained knowledge by analyzing the data does not constitute a threat to the user. However, attackers can easily break this method by monitoring the IP address or linking all data to a user's nickname. If the user's locations are always revealed as in LBS, his real identity can be inferred also [26].

B. ENCRYPTION

This method is used if there is trust between the user and the SP so that the two parties agree on a shared key to encrypt the data, and the malicious party cannot view it. However, as mentioned previously, the trust between the SP and the user may not be required in many modern applications like crowdsourcing based. Furthermore, the SP may be a malicious party seeking to collect data about users and breach their privacy. Generally, this method can be used when the user trusts the SP [27].

C. DATA SUMMARIZATION

This method reduces the amount of data sent to the server, which means the data is summarized. For example, sending the average consumption during a period for all devices is better than sending the rate of consumption of each electrical device in the smart house (from the privacy perspective) [28]. Therefore, the SP cannot reveal additional information about the user's life in this way. Contrarily, protection methods based on data summarization use Data Mining (DM) algorithms to find important information from the data before sending it, then only send the information that the SP needs without additional data to be analyzed, which causes discovering sensitive data about the user as time passes [29]. However, for crowdsourcing-based services, which need as much detailed data in real-time as possible, this protection method is not acceptable.

D. ACCESS PERMISSION

Many SPs are now relying on this method to notify the user that they care about their privacy, where the user has the right to access his data anytime and anywhere, in addition to the right to modify and delete the data. However, if the SP is malicious, it can easily make a copy from data in another place. Therefore, this method is not enough to protect privacy from this type of attack. This method can enhance the privacy and security of an external attacker [30].

E. TRUSTED THIRD PARTY (TTP)

Sometimes the SP cannot be trusted, so a third party can be a broker between the user and the SP. The third-party isolates the users' identities from the SP, and the SP cannot obtain information about the users. However, the third party may pose a threat to the privacy of the user if it is hacked or if it is malicious. If this approach is enhanced, it can be a good option for some crowdsourcing-based services [31].

F. OBFUSCATION

This approach depends on adding noise to the data before sending it to the SP or making an amendment to some

TABLE 1. Comparison of the privacy techniques.

Approach	Method	Protected Part	Drawbacks	Fit Crowdsourcing	Main Attacks
Anonymity [26]	Nickname	ID	Very Weak	No	Linking Data + Data reliability
Encryption [27]	Cryptography	Data / Query	Trust to SP	No	Malicious SP
Data Summarize [28]	DM and Statistics	Data / Query	Delay	No	Tracking Data
Access Permission [30	Authorization	Data / Query	Trust to SP	No	Malicious SP
TTP [31]	Trusted Broker	Data / Query	Trust to TTP	Can be	Malicious TTP
Obfuscation [32]	Add noise	Data / Location	Accuracy of Result	No	Path Tracking + Homogeneity
Dummy [33]	Send false queries	Query / Location	Accuracy of Result	No	Map Knowledge + Homogeneity
Collaboration [35]	Peers Cooperation	Query / Location	Trust to Peers	No	Homogeneity
Cloak Area [36]	Anonymizers	Query / Location	Trust to Anonymizer	Can be	Areas Tracking
PIR [37]	Query huge data	Query / Location	Load and Resources	No	DOS

data to prevent the SP from obtaining accurate information. However, some of these processes are not acceptable in crowdsourcing-based services which require accurate locations (like health or transportation services). This method is good at protecting the privacy of the user's location, which is one of the most dangerous parts of information that a malicious attacker can exploit. Some methods of obfuscation send a nearby location or a landmark in the same area instead of the user's exact location. However, in dynamic queries (frequently contacting SP), this approach can be hacked, where an attacker can draw a path for the user's movement, then predict the location of his presence in a particular area at a certain time. Furthermore, if the obfuscation area is homogeneous (for example, all the buildings are related to medical activities), the attacker can obtain true information about all users in the cell without needing an accurate location [32].

G. DUMMY

This approach depends on sending many false queries to the SP alongside the real one so that the SP cannot differentiate between them. Therefore, if this data is stored with the SP side and analyzed, the SP will get misleading information about users and their interests. Thus, this method seeks to protect the privacy of transmitted data as location data. But it may not be suitable for many crowdsourcing-based services like transportation services (traffic issues), in which the number of inquiries or vehicles located in a particular area has to be accurate. Also, generating a smart dummy is difficult to be detected, especially with moving users (dynamic queries) [33], [34].

H. COLLABORATION AMONG PEERS

This is one of the good ways to protect privacy if the SP poses a threat to users. Therefore, the goal is to reduce the amount of data that can be collected by the SP. To achieve that the users exchange the results of their queries. For example, user A can inquire about a specific target from B or C who are in the same area instead of calling the SP. There are many other ways to collaborate. In general, this method needs to have many users in the same area to be somewhat effective, as it sometimes causes delays compared to communicating directly with the SP [35].

I. CLOAK AREA

Cloak Area is the development of the TTP approach by dividing the area into many regions or cells. Each cell has an Anonymizer that protects the users' privacy from SP. Anonymizer will manage the peers inside its region only. Therefore, it will not pose a threat to the users' privacy when users deal with Anonymizer within their cells. This approach can be accomplished in another way through the cooperation of users to send all the queries at once by one of them to mislead the SP. That will prevent the SP from collecting information about each user, known as K-Anonymity [36].

J. PRIVATE INFORMATION RETRIEVAL (PIR)

The user requests a large amount of data from the SP, then stores it in his memory. For future queries, the user will search in his memory without the need to contact the SP for each query. That means the SP cannot determine what the user wants. But this method requires considerable capabilities and resources for the user, which may not be available on many devices [37].

K. HYBRID TECHNIQUES

These methods rely on integrating and merging techniques to provide more security and privacy for better performance. For instance, users can use a cache of TTP to reduce the need to contact the SP and improve system performance. Additionally, some methods use cooperation between users to create smart dummies.

In mobile devices, manufacturing companies have started implementing many policies to enhance their users' privacy [38]. For example, in the IOS operating system, the user can review all the permissions that applications require such as permission to access contact information, messages, media, camera, or mic. Furthermore, when using the mic or

the camera, an indicator will appear to notify the user that the camera or mic is currently being used by an application, even if it is a hidden usage. Also, some permissions can be given only at the time of using the application by the user [39], [40]. Moreover, the user can select the option to prevent access to his accurate location if he gives the location permission to an application. All the previous options mentioned are also good and useful, but still not enough to preserve the users ' privacy from the SP like the iPhone company. Therefore, privacy still needs more effective compatible solutions with different applications.

1) SUMMARY

All previous approaches and methods of protection have drawbacks. These draw-backs prevent those protection techniques from being suitable for crowdsourcing-based services, which require enhanced methods. Table 1 summarizes the previous techniques and their drawbacks in addition to the attacks of each approach. Most traditional approaches do not suit crowdsourcing-based services, which require accurate data without delay. Even the anonymity approach is very weak in the protection perspective. The TTP and Cloak-Area can be enhanced to suit crowdsourcing-based services.

Previously, we have presented our hybrid approach, which depends on the doubling of protection to improve privacy and security. We presented Double Cache (DCA) [41] and Double Obfuscation (DOA) methods [42]. Nonetheless, neither DCA nor DOA are suitable for crowdsourcing-based services too, as this kind of service requires maintaining the data accuracy without large delay in processing. Thus, in this research, we present an enhanced approach called ''Double Cloak Area'' (DCL-Ar) which is suitable for crowdsourcing-based services by providing accurate information, a higher level of protection, and reliable data, without having a significant effect on the performance.

Taxonomy Table for Privacy Protection Methods in Crowdsourcing-based Services.

In the core section of our research (Section III), we provide a comprehensive exploration into the DCL-Ar mechanism, including its design, comparative analysis with existing methodologies, and detailed empirical findings. We delve into the innovative use of fog computing within DCL-Ar to augment privacy protection, highlighting its architecture as illustrated in Figure 4 and its unique approach in managing user privacy through advanced techniques. This analysis extends across various sectors such as health, transportation, business, and social media, showcasing DCL-Ar's critical role in enhancing privacy in crowdsourcing services and mitigating inherent privacy threats. By ensuring data accuracy and user anonymity, DCL-Ar addresses the vulnerabilities of traditional methods, supporting the integrity and confidentiality of user data in diverse applications. Through rigorous simulations, we demonstrate DCL-Ar's superiority in privacy protection, operational efficiency, and data integrity, thereby solidifying its position as a significant advancement in smart city privacy preservation.

The DCL-Ar approach, along with SPF and DOA, achieves maximum privacy (Entropy = 1) by avoiding direct user-SP communication, which contrasts with the limited privacy in dummy and traditional Cloak-Area methods where user identities are more vulnerable. DCL-Ar's dual-level anonymization significantly enhances user privacy by blending queries and locations within a vast user pool, achieving unparalleled privacy standards. It also leads in minimizing query transmissions to the SP, thereby reducing the risk of information exposure and optimizing system performance. This efficiency is further evidenced by DCL-Ar's superior cache utilization, which notably improves hit rates, especially beneficial as users navigate through different areas. DCL-Ar outperforms other methods in response speed by eliminating the need for extensive data reprocessing and leveraging efficient cache strategies, which not only expedite responses but also minimize update times by selectively refreshing duplicated query positions. Ultimately, DCL-Ar demonstrates a significant reduction in total response time for queries, especially at higher cache hit rates, underscoring its efficacy in privacy preservation while maintaining system efficiency.

III. GAP ANALYSIS

All smart services and systems fundamentally rely on timely data availability for their effectiveness. Crowdsourcing plays a crucial role in providing real-time data directly from the heart of events, a process facilitated by the widespread availability of smart devices and communication networks. However, implementing crowdsourcing efficiently within services faces two significant challenges.

A. FIRST CHALLENGE: PRIVACY CONCERNS

The primary challenge is the privacy of data provided by users engaged in crowdsourcing. Privacy concerns deter many individuals from sharing their data, potentially leading to the failure of crowdsourcing initiatives. This issue is particularly acute when the data includes sensitive personal information such as user locations. Without robust privacy protection, this data can be exploited, leading to the tracking of individuals and unauthorized disclosure of personal details such as habits, religious beliefs, social status, workplaces, and more.

B. SECOND CHALLENGE: DATA RELIABILITY

The second challenge concerns the reliability of the data provided by users and the need to minimize false or anomalous information. Such inaccuracies significantly impact the quality of crowdsourcing data and, consequently, the performance of dependent services. The existing gap in current privacy protection methods, which are illsuited for crowdsourcing, compounds this issue. Traditional methods often involve altering the core data to protect privacy, such as obfuscation or K-Anonymity, or adding spurious data to genuine data (e.g., Dummy data). These approaches can render the concept of crowdsourcing ineffective.

Moreover, other protection methods like Private Information Retrieval (PIR), Trusted Third Parties (TTP), Anonymity,

or encryption can introduce unacceptable time delays or provide inadequate protection, thus significantly impacting performance.

C. PROPOSED SOLUTION: DCL-AR APPROACH

There exists a notable gap in traditional privacy protection mechanisms, making them unsuitable for services that rely on crowdsourcing. This gap necessitates the development of new protection methods tailored to meet the specific needs of these services and applications. The research proposes a novel approach called DCl-Ar, designed to safeguard privacy while enhancing the reliability of crowdsourcing data. This approach addresses both the privacy concerns and the need for reliable data without compromising the operational efficacy of crowdsourcing platforms.

IV. THE PROPOSED METHOD–DOUBLE CLOAK AREA (DCL-AR)

This Section unveils DCL-Ar, an advanced technique building upon the traditional Cloak-Area method to bolster privacy in crowdsourcing applications. It acknowledges limitations in existing methods, where location anonymization and user movement can hinder service accuracy. DCL-Ar tackles this by extending the cloak area through collaboration between users and fog nodes, safeguarding both data accuracy and user privacy. The approach is presented through three illustrative scenarios, showcasing its flexibility and effectiveness in preserving privacy without sacrificing system performance or data integrity, thus addressing crucial challenges in location-based services and crowdsourcing.

The Cloak-Area approach is an approach that can be developed to be suitable for crowdsourcing applications with better security. In the traditional Cloak-area method, all users in the area send their queries to the Anonymizer. The Anonymizer replaces exact queries' locations with a unified location. The Anonymizer then sends the queries on behalf of peers (users) to the SP to hide their identity. Location replacement is necessary to protect the privacy of user locations because both the cloak area and the number of users are small. Moreover, in the case of dynamic queries (users are in constant motion), the malicious SP may be able to isolate the queries, track their users, and reveal their identity later. The SP can also implement an area tracking attack, and it can also implement a homogeneity attack as well as reveal additional information about users if all Points of Interest (PoI) in the cloak area are of the same nature.

Unfortunately, the process of changing users' locations causes additional weaknesses in the traditional Cloak-area approach. We can summarize them in the following points:

- Negative impact on the accuracy of the basic service. This approach will not suit precise LBS like many Crowdsourcing services.
- The Anonymizer needs to reprocess the data returned from the SP (in the case of queries) to map real user locations, and this negatively affects performance.
- The user may leave the region for another region before receiving the results of his query from the Anonymizer.

To solve the previous problems and challenges, it has to provide a developed approach that ensures the accuracy of data and queries' locations and ensure adequate protection for users from tracking. Therefore, there is a need to introduce a solution to the homogeneity in the area and the small number of users. This paper presents the idea of the ''DCL-Ar'' enhanced approach. This approach ensures the formation of a large cloak area with a lot of users and without overhead affecting the system's performance on the Anonymizer. Moreover, the ''DCL-Ar'' approach solves other challenges and weaknesses, such as when the user leaves the cloak area before receiving the result. Additionally, the approach supports working with crowdsourcing services that require accurate data.

The main idea of the proposed approach is to multiply and expand forming a first private cloak area by peers then second one by fog nodes to form a larger and more protection area. In other words, the DCL-Ar depends on two cloak areas. The Cloak-area1 is among peers, and the fog node in each one plays the role of Anonymiz-er1. The second area, Cloak-Area2, is among adjacent fog nodes (by Anonymizers themselves). Creating the Cloak-Area2 has three different scenarios for collaboration and selection of Anonymizer2 (Manager of Cloak-Area2).

In the first scenario, the Core-Fog node is Anonymizer2 if a layer for core-fog nodes is available. In the second scenario, one of the cooperating fog nodes is selected to act as Master Anonymizer2 on each communication with the SP. In the last scenario, all fog nodes send data to the SP together, which means that each fog node is playing two roles: Anonymizer1 and Anonymizer2. The three scenarios help the proposed approach to be dynamic in working with the available architecture and diversity of Crowdsourcing applications and services. Figure 3 depicts the high-level architecture of the DCL-Ar.

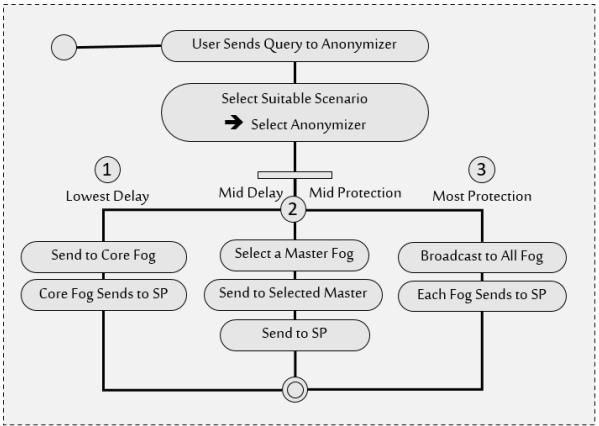

FIGURE 3. High level architecture of DCL-Ar.

As depicted in Figure 4, our innovative approach unfolds three distinct sections, each illuminating a specific scenario addressed by our methodology. The left segment of the

figure reveals a structured layout with two layers of fog nodes. Positioned at the top is the core fog (Anonymizer 2), overseeing a cluster of fog nodes below. The subsequent layer, known as the fog layer, encompasses multiple fog nodes (Anonymizer 1), each managing a group of users, referred to as a cloak. In this scenario (Figure 4.A), users, upon connecting to a new fog node, select a new nickname.

Following that, users elegantly transmit data or queries, alongside their precise locations, to the designated fog node. The fog node meticulously examines the data for issues and identifies unreliable data sources. Subsequently, it compiles received queries without altering their locations and forwards this data to the Core-Fog. The Core-Fog, in turn, sends all query sets to the Service Provider (SP) without revealing user or fog node IDs. The SP receives the data or queries from the Core-Fog but remains unaware of user or fog node specifics. After processing the information, the SP returns results to Core-Fog. Anonymizer 2 then disseminates the copies of result to individual fog nodes, each of which conveys its results to users in its designated area. For queries originating from other regions, fog nodes store them in their cache. This enables responsive handling of users' inquiries from nearby areas without the need to connect to the SP.

In the middle section (Figure 4.B), it is evident that the core fog has been omitted. The structure now comprises a group of fog nodes responsible for interfacing with the Service Provider (SP). Initially, one of the fog nodes is designated or elected as the master node (Anonymizer 2). Like the previous scenario, when any node (user) enters a cell, an alias is chosen, and subsequently, it sends its data and requests to a specific fog node (Anonymizer 1). This node then forwards the information to the previously designated master node, which in turn, relays it to the SP. Conversely, information from the SP is directed to the primary fog node, which then disseminates it to other fog nodes, ultimately reaching the users. After that, fog nodes elect another node as the master one.

In the third section (Figure 4.C), a similar method is employed. Each user selects a nickname upon entering a designated area associated with a fog node. The user transmits data and requests to the fog node, which collects the requests without altering them. Each node exchanges the compiled requests with its neighboring nodes and subsequently each node transmits the aggregated information from different areas to the SP. In contrast, the SP returns the results to the nodes communicating directly with it, which in turn, relay the results to neighboring nodes. The final dissemination reaches users within their respective areas.

A. COMPARISON BETWEEN THE THREE SCENARIOS

Table 2 compares three scenarios, with scenario 1 (A) being the simplest but least efficient in terms of both processing and privacy protection. It employs a static Anonymizer (Core Fog), which may lead to bottlenecks. It offers a moderate privacy as the core fog itself can pose threat and identify users and their locations. Scenario 2 (B) introduces a master fog node to enhance privacy by collecting and forwarding data from all fog nodes in a given area. This reduces the load on individual fog nodes and providing better privacy protection than scenario 1. Scenario 3 (C) further improves efficiency of privacy preserving where all fog nodes send similar data to mislead the SP. It is reducing the reliance on the service provider. This scenario offers very good privacy protection, as the SP cannot identify users or fog nodes, however, it is also the most complex to implement and cause higher overload.

In summary, scenario 3 is the most efficient in the privacy perspective, but it is worst in performance and complex to implement. Scenario 2 strikes a good compromise between performance and privacy, making it suitable for most applications. Scenario 1, while the simplest to implement with better performance, but it is the least privacy protection and less availability (single point of frailer).

B. MAIN ALGORITHM OF DCL-AR

```
//User or Peer
Result = Send (Anonymizer1, Nickname, Location, Data);
//Anonymizer1
List_Queries1 [] = null;
List_Queries2 [] = null;
While (true)
  New_Query = Get_Query();
  List_Queries1.add (New_Query);
  List_Queries2.add (Anonymizer1,
  New_Query.Location, New_Query.Data)
Anonymizer2 = Find_Anonymizer2();
// Scenario1 for Find_ Anonymizer2
Anonymizer2 = get_Core_Fog (Anonymizer1.Location)
Results = Send (Anonymizer2, List_Queries);
// Scenario2 for Find_ Anonymizer2
Counter = Get_Next_Counter();
List_Fog = Get_Neighbors(Anonymizer1.Location);
Anonymizer2 = Counter mod List_Fog.Count;
Results = Send (Anonymizer2, List_Queries);
// Scenario3 for Find_ Anonymizer2
List_Fog = Get_Neighbors(Anonymizer1.Location);
For (int i =1; i<= List_Fog.Count; i ++)
    Results = Send (List_Fog[i]; List_Queries);
List_Queries.Clear();
If (Anonymizer1.ID == Anonymizer2)
 While (true)
          List_Queries [] = Get_List ();
          List_Queries2.Add (List_Queries );
          Results = Send (SP, List_Queries2);
          List_Queries2.clear();
          Return (List_Queries.Anonymizer, Results);
For (int i =1; i<= Results.Count; i ++)
   Return (List_Queries1[i].UserID, Results[i]);
End;
//Anonymizer2
While (true)
   List_Queries [] = Get_List ();
   Results = Send (SP, List_Queries);
Return Result
```

The above algorithm is designed for managing user requests and data in a distributed fog computing environment, featuring three main components:

TABLE 2. Comparison among the proposed scenario.

Feature	Scenario 1-A	Scenario 2-B	Scenario 3-C
Anonymizer Selection	Static	Rotating	Multi-hop
Data Sharing	Limited	Limited	Extensive
Query Processing	Centralized	Centralized	Distributed
Performance	High	Moderate	Low
Privacy	Moderate	Good	Very good

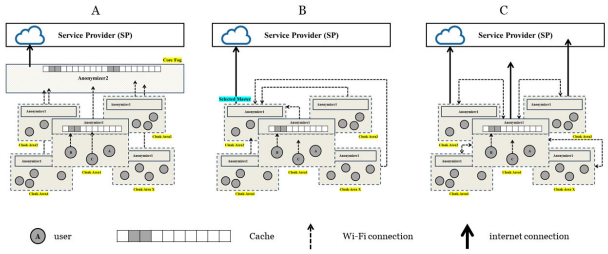

FIGURE 4. Scenarios of DCL-Ar approach.

User or Peer: Represents users or peers submitting data or queries along with their location and nickname to Anonymizer1.

Anonymizer1: The first anonymizer in the fog network, storing information in List_Queries1 and List_Queries2. Periodically, Anonymizer1 retrieves a fresh question from List_Queries1 and forwards it, along with location and contents, to Anonymiz-er2.

Anonymizer2: The second anonymizer in the fog network, forwarding queries from Anonymizer1 to the service provider (SP) for handling. Results from the SP are received by Anonymizer2 and then sent back to Anonymizer1.

The algorithm includes three scenarios for finding Anonymizer2:

Scenario1: Using the get_Core_Fog function to find Anonymizer2 based on Anonymizer1's location.

Scenario2: Using a round-robin approach to select the next Anonymizer2 in the list of neighboring fog nodes based on a counter.

Scenario3: Retrieving a list of neighboring fog nodes, sending queries to each, and combining results.

Anonymizer1 continuously pulls lists of queries from List_Queries2 and sends them to the SP if Anonymizer1 and Anonymizer2 are the same. After processing, Anonymizer1 receives answers from the SP, associates user IDs with questions from List_Queries1, and combines them with results. Ultimately, users receive a combined list of user IDs and results from Anonymizer1. In summary, this algorithm provides a distributed and anonymous method for handling queries and data in a fog computing environment.

V. REAL EXAMPLES ABOUT CROWDSOURCING SERVICES AND PRIVACY THREAT

This section reviews some services and applications of the main areas, how these services can be exploited to penetrate the privacy of users, and then how the proposed approach can protect their privacy.

A. HEALTH

Many concepts have developed after the emergence of the IoT, starting from Smart Health and Ubiquity Health to the Internet of Healthy Things (IoHT). These technologies depend mainly on providing vital measurements and data about the patient in real-time, as this data is collected from sensors, whether wearable or spread in the environment surrounding the user. This data is processed and analyzed for early detection of any serious or potential change in the user's health. Recently, with the Corona crisis, the number of digital services in the health field increased, and homes turned into health centers based on IoT [43].

Moreover, many services and health centers have depended on crowdsourcing to get more information about users during the pandemic. Unfortunately, the malicious SPs have enabled us to collect a lot of sensitive data and violate users' privacy, for example, by tracking users' activities and locations [18]. Thus, the question was, how is it possible to work with medical pandemics and collect useful data, without jeopardizing user privacy and threatening it? DCL-Ar can be used with most health services that don't require a specific identity for the user. Moreover, DCL-Ar can utilize the anonymizer node to compare data and detect any abnormal or unreliable data.

B. TRANSPORTATION AND TRAFFIC MANAGEMENT

Traffic systems have also evolved greatly following the IoT revolution. Smart vehicles, smart traffic lights, digital streets, street conditions, smart parking lots, as well as various LBS have appeared. Most of these services depend on crowdsourcing. Malicious parties or providers of mapping or location-based services can track a user's location for a period to discover considerable user data which is not associated with the advertised service [44].

For instance, when analyzing the places that the user visits in a month, it is possible to know the nature of the user; where his home is, when he leaves the house, if he is sick and visits health centers, if a visit is to luxury or popular restaurants, if an employee is active or if he arrives on time, how often is the employee late, if he has children in school or not, and many other data and sensitive data that SP is not allowed to

access. DCL-Ar does not allow SPs to link data to a specific user or create a profile for any user who uses these services. Also, DCL-Ar utilizes the anonymizer to detect and isolate anomalous data.

C. BUSINESS

Electronic invoices and purchases online are significantly diffused, by analyzing invoices that can reveal data about user behavior, wealth or spending rate, income, and when there are special occasions in his life. Moreover, the delivery services reveal the user's location and places (at home or work). Recently, many companies utilized crowdsourcing data and tools like reviews, comments, questioners, etc., to get much information about their products and customers. Unfortunately, a huge amount of the collected information is private, and users do not know about it [45]. DCL-Ar enables companies to get purposive information about their products and customers without disclosing the customers' identities.

D. SMART PHONE AND SOCIAL MEDIA

Smartphones have evolved into the epitome of intelligent devices intricately woven into users' daily existence. Their sophistication lies in the myriad services and applications they encompass. Notably, social networking applications have ascended to prominence, becoming integral facets of users' lives. These platforms serve as conduits for expressing opinions, showcasing talents and hobbies, fostering discussions, and functioning as primary outlets for news consumption [46]. Smartphones and social media are the main tools for most crowdsourcing services. Thus, at the same time, they are a main threat to the user's privacy which can reveal his interests, behavior, hobbies, preferences, inclinations, and a lot about his private life [47]. Although smart services require more and more user data to provide continuous enhancement and better quality, privacy must be guaranteed. DCL-Ar can help in this part.

E. SUMMARY

Briefly, after studying some real privacy threats and penetration, we realize the danger of modern applications and technologies that are spreading all around us like many services based on crowdsourcing. The proposed approach can play an important role in supporting these services and their future by preserving the user's privacy in addition to insuring the reliability. The next section will discuss and prove the effectiveness of DCL-Ar in protecting privacy without.

VI. RESULTS AND DISCUSSION

To prove the superiority and efficiency of the proposed "DCL-Ar" approach, this section presents the results of comparison with several other methods, namely.

- Without using any protection method.
- Using the traditional Cloak-Area approach (Cloak Area) [36].
- Using the obfuscation approach (DOA) [42].
- Using the peer-to-peer approach (SPF) [48].
- Using the dummy approach (Enhanced-CaDSA) [38].

The comparison relied on standards to measure the level of privacy and protection achieved and the level of affecting performance. The main standards are:

✓ K-Anonymity [24], [48] It is expressed by dividing 1 by the number of queries sent to the SP, whether true or false.

$$K - Anonymity = \frac{1}{(1+x)}$$

where K = number of false queries

✓ Entropy [24], [48], which represents the amount of real information that the SP can link to the sender's user, or the percentage of the SP's certainty that the information it has, is related to a specific user.

$$Entropy = -\sum_{i=1}^{n} P_i * \log_2 P_i$$

where n is number of sent queries, and pi is probability of query i belongs to the user.

There is another standard for privacy, but it is completely related to entropy, for example, Estimation Error, which is a percentage of Entropy, and the Ubiquity standard, that is, the spread of the user everywhere within the region, which is Entropy^2.

✓ Time is related from the moment the query is sent to the moment the result is received by two main factors: Send Time (ST) and Process Time (PT). Time is also affected by the data size (DS), the number of queries sent (NQ), and the Cache Hit Ratio (CHR) [24], [48].

✓ There are non-quantitative standards related to the level of efficiency of the proposed approach and its robustness in the face of attacks, in addition to its ability to adapt to specific applications and services such as Crowdsourcing.

A. MATHEMATICAL ANALYSIS

Based on the previous criteria and standards, each of the compared methods was analyzed.

Table 3 contrasts how well various privacy-preserving techniques perform for location-based searches. Seven criteria are used to assess the methods: k-anonymity, entropy, cache, duration, data size, crowdsourcing appropriateness, and potential attacks.

1) COMPARISON BASED K-ANONYMITY

- With a k-anonymity of 100%, the suggested strategy, DCL-Ar, achieves the highest k-anonymity value of all the approaches. This implies that there are at least 100 other users in the same anonymization group for each query made by a user, making it challenging to pinpoint the exact user who submitted the query.

TABLE 3. Comparison among main privacy methods.

Method	K-Anonymity	Entropy	Data Size	Time	Cache	Suitability for Crowdsourcing	Potential Attacks
Without Protection	1	0	1 KB	ST	Yes	No	-
Traditional Cloak-Area	1/(1+K)	-1/k * Log(1/k)	1 KB	ST + T_anonymizer + T_mapping	Yes	Limited	Homogeneity, tracking area
Dummy Approach (Enhanced-CaDSA)	1/(1+K)	-1/k * Log(1/k)	(1+K) KB	(1+K) * ST	Limited	No	Tracking real query, detecting dummies
Obfuscation Approach (DOA)	1/(1+D)	-1/D * Log(1/D)	1 KB + obfuscation area range	ST + T_mapping	Limited	No	Homogeneity, tracking area, map knowledge
Peers-Cooperation (SPF)	Maximum	1	1 KB	ST + T_swapping + T_fog + T_fog_peers	Yes	Medium	Peer attacks, cooperated peer disconnect
Proposed Approach (DCL-Ar)	Maximum	1	1 KB	ST + T_anonymizer1 + T_anonymizer2	Yes	High	Fog cooperation with SP

- The Traditional Cloak-Area method achieves a k-anonymity value that is inversely proportional to the number of cooperated users in the area. This means that the k-anonymity value can vary depending on the specific location of the query.
- The Dummy Approach (Enhanced-CaDSA) and Obfuscation Approach (DOA) both achieve a k-anonymity value that is inversely proportional to the number of dummies or false locations in the area, respectively. This means that the k-anonymity value can vary depending on the specific implementation of the method.
- The Peers-Cooperation (SPF) method achieves a maximum k-anonymity value, but it is only applicable in certain scenarios where users are willing to swap their query locations with other users.
- The Without Protection method provides no k-anonymity, as the user's real location is directly sent to the service provider.

2) COMPARISON BASED ENTROPY

- The proposed approach, DCL-Ar, achieves the highest entropy value of all the methods, with an entropy of 1.0. This means that there is complete uncertainty about the location of the user who made the query.
- The Traditional Cloak-Area, Dummy Approach (Enhanced-CaDSA), and Obfuscation Approach (DOA) all achieve entropy values that are less than 1.0, indicating some degree of uncertainty about the user's location. The exact entropy value for each method depends on the specific implementation.
- The Peers-Cooperation (SPF) method achieves a maximum entropy value of 1.0, but it is only applicable in certain scenarios where users are willing to swap their query locations with other users.
- The Without Protection method provides no entropy, as the user's real location is directly sent to the service provider.

3) COMPARISON BASED DATA SIZE

- The Traditional Cloak-Area and Peers-Cooperation methods have the smallest data size of all the methods, at 1 KB per query. This is because these methods only send the query location and a small amount of additional information to the service provider.
- The Without Protection, Dummy Approach (Enhanced-CaDSA), Obfuscation Approach (DOA), and Proposed Approach (DCL-Ar) all have a data size of 1 KB per query, plus some additional data depending on the specific implementation of the method.
- The additional data required for the Dummy Approach (Enhanced-CaDSA) and Obfuscation Approach (DOA) is used to store dummy locations or false locations, respectively. The additional data required for the Proposed Approach (DCL-Ar) is used to store anonymization information.

4) COMPARISON BASED TIME

- The Without Protection method has the shortest processing time of all the methods, at ST (where ST is the time it takes for the service provider to process the query). This is because the Without Protection method does not require any additional processing steps.
- The Traditional Cloak-Area method has a processing time of ST + T_anonymizer, where T_anonymizer is the time it takes to anonymize the query location.
- The Dummy Approach (Enhanced-CaDSA) and Obfuscation Approach (DOA) both have processing times of (1 + K) * ST and ST + T_mapping, respectively where K is the number of dummies or false locations and T_mapping is the time it takes to map the query location to a dummy or false location.
- The Peers-Cooperation (SPF) method has a processing time of ST + T_swapping + T_fog + T_fog_peers, where T_swapping is the time, it takes to swap query

locations with other users, T_fog is the time it takes to send the query to the fog layer, and T_fog_peers is the time it takes to swap query locations with other fog nodes.
- The Proposed Approach (DCL-Ar) has a processing time of ST + T_anonymizer1 + T_anonymizer2, where T_anonymizer1 and T_anonymizer2 are the times, it takes for the two anonymizers to process the query.

5) COMPARISON BASED CACHE

- The Without Protection, Traditional Cloak-Area, and Peers-Cooperation methods all support caching. This means that the query results can be stored so that they do not need to be recomputed each time they are requested.
- The proposed approach, DCL-Ar, also supports caching, and in addition, it can store data from neighbors' caches, further improving performance.

6) COMPARISON BASED SUITABILITY FOR CROWDSOURCING

- The proposed approach, DCL-Ar, is the most suitable method for crowdsourcing applications because it has the highest k-anonymity and entropy values, the smallest data size, the shortest processing time, and support for caching.
- The Traditional Cloak-Area, Dummy Approach (Enhanced-CaDSA), and Obfuscation Approach (DOA) are all less suitable for crowdsourcing applications because they have lower k-anonymity and entropy values, larger data sizes, and longer processing times.
- The Peers-Cooperation (SPF) method is also less suitable for crowdsourcing applications because it requires users to be willing to swap their query locations with other users, which may not always be feasible.
- The Without Protection method is not suitable for crowdsourcing applications because it provides no privacy protection.

7) COMPARISON BASED POTENTIAL ATTACKS

- The proposed approach, DCL-Ar, is resistant to most known attacks, including homogeneity attacks, tracking area attacks, and map knowledge attacks. However, it is still possible for a malicious fog node to collude with the service provider to de-anonymize user locations.
- The Traditional Cloak-Area method is susceptible to homogeneity attacks and tracking area attacks.
- The Dummy Approach (Enhanced-CaDSA) is susceptible to tracking real queries and detecting dummies attacks.
- The Obfuscation Approach (DOA) is susceptible to homogeneity attacks, tracking area attacks, and map knowledge attacks.
- The Peers-Cooperation (SPF) method is susceptible to peer attacks and cooperated peer disconnected attacks.
- The Without Protection method is susceptible to all known attacks.

Briefly, the proposed approach, DCL-Ar, provides the best overall performance for privacy-preserving location-based queries. It achieves the highest k-anonymity and entropy values, has a small data size and a short processing time, and is suitable for use in crowdsourcing applications. However, it is still possible for a malicious fog node to collude with the service provider to de-anonymize user locations.

B. EXPERIMENT CONFIGURATION AND DATA

During the configuration of our work experiment, we employ a methodical approach, and replicate the data framework and assumptions utilized in preceding scholarly works. This strategy facilitates the generation of comparative results, subsequently illustrated in detailed figures. The parameters defining the simulation environment are meticulously established, encompassing several key aspects:

1. Query Size: Each unprotected query within the simulation is standardized at 1KB, ensuring uniformity in data handling and processing.
2. User Base: The simulation is scaled to accommodate 10,000 users, offering a robust model for user interaction and system load.
3. Geographical Regions: The simulated landscape is divided into 10,000 (100 × 100) discrete regions, providing a comprehensive spatial framework for data analysis.
4. Cache Capacity: A cache size of 100KB is designated, challenging the system's ability to efficiently manage and retrieve data.
5. Points of Interest (POIs): 100 POIs are strategically embedded within the simulation, serving as focal nodes for user queries and interactions.

The technical execution of the simulation is underpinned by the use of C# as the primary programming language, chosen for its robustness and versatility. Data management and storage are handled through SQL Server databases, a decision that underscores the emphasis on high-level data integrity and accessibility. The entire simulation framework is developed and operated within the environment of Visual Studio.Net 2019, ensuring a cutting-edge technological foundation.

A critical component of the input of simulation's data is derived from the Geo-Life dataset. This dataset, encompasses 17,000 tracks from 180 users collaboratively monitored over a three-year period, provided a real-world context to the simulation. The inclusion of this dataset not only enhances the simulation's relevance and applicability but also aids in benchmarking its performance against realistic user behavior patterns and movement trajectories. The comprehensive nature of this dataset, coupled with the rigorously defined simulation parameters, positions the experiment as a significant contribution to the field, offering insights into user-data interaction within a controlled yet realistic digital environment.

C. RESULTS AND DISCUSSION

Figure 5 shows that the proposed approach, in addition to the SPF and DOA approach, achieves privacy protection with a maximum Entropy (Entropy = 1) due to the user not communicating with the SP directly, or the user sending a query to another user as in SPF. In the dummy approach, the SP can obtain part of the user's information within the set of sent dummy-queries. Also, in the traditional Cloak-Area, the SP can detect the identity of users and can reveal information about them in the case of a small area or a small number of users.

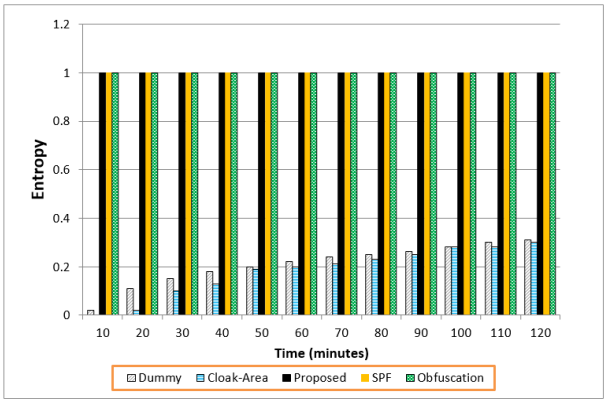

FIGURE 5. Comparison based on entropy.

Moreover, the proposed protection method (DCL-Ar) is distinguished by using two levels to hide the user's identity from the SP (Anonymizer1 and Anonymizer2), in addition to hiding the user's query and data within a large number of users and hiding user's location within a large area with a large number of PoI. Thus, the proposed approach guarantees a high level of privacy. And since the achieved privacy is maximum (Entropy = 1), then both standards (Ubiquity and Estimation Error) will be Maximum as well, because these two standards are related to Entropy. This is logical in the DCL-Ar because users' queries are not modified, and the large users' number means that they are spread in all places within the cloak area, and thus the probability that the SP will be wrong in specifying the user's location is maximum too.

Figure 6 shows the number of queries sent to the server, which has a role in determining the level of system performance and the level of privacy as well. As the number of queries sent to the SP is lower, the information is less likely to be disclosed, and there will be less load on the system, and a better performance rate can be achieved. In terms of the number of queries, the type of protection used plays a big role, in addition to the use of cache memory in the area where the users are located. The results show the superiority of the proposed approach DCL-Ar in addition to the SPF in achieving the lowest number of sent queries. The justification for this is that both approaches do not use dummies to increase the number of queries on one hand, and do not modify the query location, which negatively affects the hit rate.

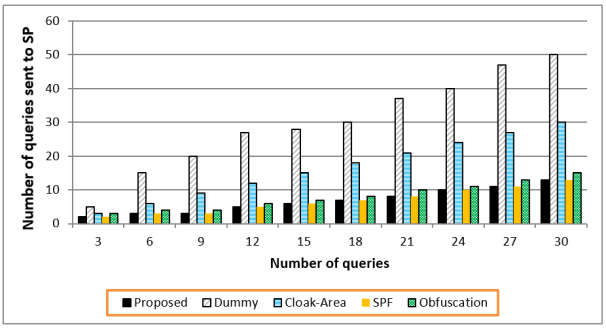

FIGURE 6. Comparison based on number of sent queries.

Moreover, both the DCL-Ar approach and the SPF use two caches, which give double value of the probability of a hit. What further distinguishes the DCL-Ar from SPF is that the cache in the proposed approach can store queries for a contiguous area. This is useful in solving the problem of moving the user from one area to another before receiving the user's result, increasing the hit rate, and benefiting more from the cache.

The foregoing discussion is reinforced by Figure 7 which shows the superiority of DC1-Ar approach in terms of hit rate. This is justified by storing only real queries for users, in addition to not modifying the query locations. In Figure 7, we assume that the traditional Cloak Area also has a cache, but the hit rate is lower because all user queries are linked to one location, which is contrary to reality. The dummy method remains the worst because dummy queries are stored in the cache, which effects the hit rate negatively, even if these dummies are chosen intelligently.

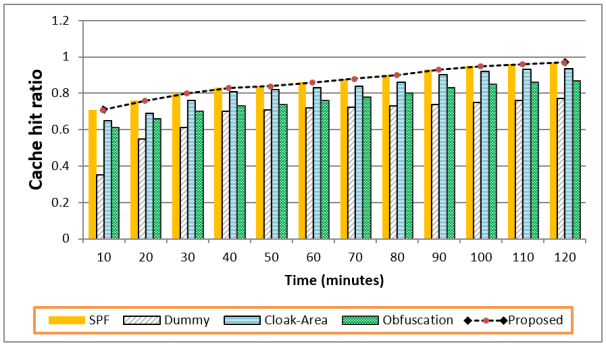

FIGURE 7. Comparison based on cache hit ratio.

Figure 8 shows the response speed standard and shows the superiority of the proposed approach DC1-Ar, followed by the SPF approach and then DOA. Cloak-Area and Dummy are the worst because of the overhead caused by dummies or the need to process the returned data in the case of the traditional Cloak-Area and mapping the results to real users' locations. The superiority of DCL-Ar is due to the lack of need to process the returned results on one hand, and the use of cache helps in the speed of the answer on the other hand, especially in the case of a high infection rate. Finally, the transmission time in DC1-Ar is less than that of SPF because

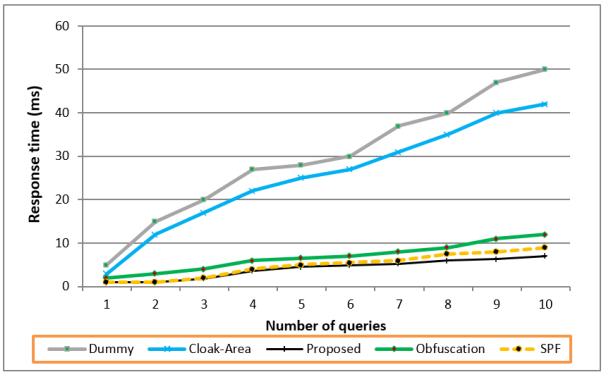

FIGURE 8. Comparison based on response time.

the proposed approach uses only two extra steps while SPF uses four steps. The additional steps are Wi-Fi connections, and therefore the transmission time is much less compared to connecting via the Internet with the SP.

The ping experiment was conducted several times for an internal connection and several times for an external connection, and then the averages were calculated. After that, the percentage difference in time was calculated, and it appears that the internal communication (locally) is 0.1 of the time required to contact an external SP.

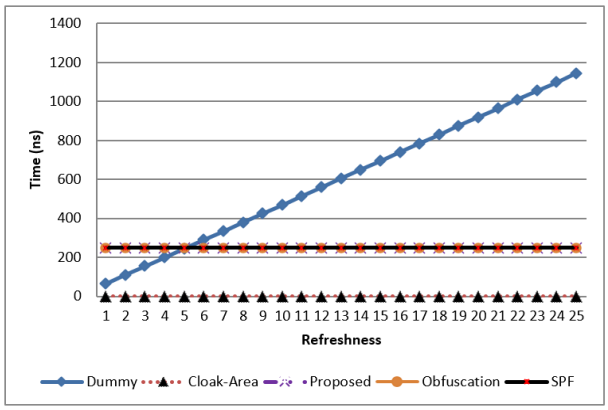

FIGURE 9. Comparison based on cache refreshness time.

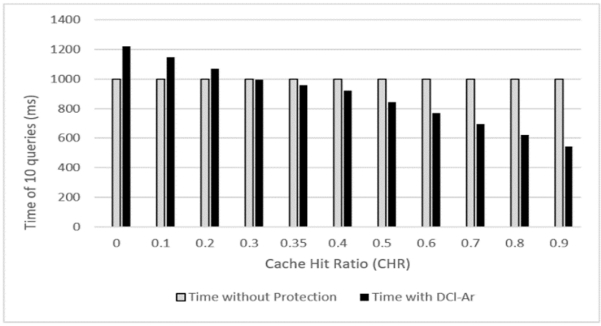

FIGURE 10. Comparison based on total time of ten queries.

Figure 9 shows the time required to update the data in the cache. If the cache is not used as in the traditional Cloak Area, there will be no time for updating. In the proposed approach DCL-Ar, the same technique is used in SPF and Obfuscation DOA is developed, where all data is constant, and only the duplicated query position is updated. However, in the dummy, all elements in the cache are swapped, so the required time increases with the number of stored queries.

Figure 10 shows the difference in the time needed to respond to ten queries between using no protection and using DCL-Ar, at different cache hit rates. It is logical that in using no protection, the time is constant, but with the protection approach that depends on cache, and when the hit rate is small, less than 3, the time is worse (greater) due to the additional protection steps. But if the hit rate is large, the proposed approach is better than without using protection and without using a cache. But note when not using protection and using a cache, time is much better, although there is no justification for using an additional party if protection such as Anonymizer is not used.

VII. DISADVANTAGES OF DCL-AR AND FUTURE TRENDS IN PRIVACY

The Double Cloak Area (DCL-Ar) methodology, while robust, is not exempt from this reality of limitations. It exhibits limitations that are outlined as follows:

1. If the set of fog nodes (Anonymizers) are malicious they can disclose the privacy of users. However, it is very rare to hack many fogs at the same time.
2. If there is cooperation between the malicious fogs and the SP, this premise is also not logical.
3. As with any protection method, there is an adverse effect on the time of processing. Nevertheless, the multi-proposed scenarios in DCL-Ar are proposed to relax previous points according to the type of services. In addition to utilize the cache in each fog node.

Furthermore, in the following points we mention some future trends in the privacy domain which we plan to work on next:

- Create a privacy protection platform like antivirus ones in the security.
- Find special protocols to protect privacy, as protocols are for protecting the data security during transmission.
- Find a tool for analyzing the level of privacy in any application like the tools of penetration testing for the security.
- Increase the awareness of users in protecting the privacy of their data and not compromising it to any party.
- Find a solution to protect privacy in pandemics, during which privacy terms and restrictions are greatly tolerated.
- Find a dynamic platform which includes many different protection methods to adapt to different IoT apps.
- Employ machine learning for smart selection from the best protection techniques for each service or application.
- Create a knowledge base for each kind of application and its related threats.

VIII. CONCLUSION

In the pursuit of advancing privacy protection within Crowdsourcing-based services and applications, this research has meticulously developed and evaluated the Double Cloak Area (DCL-Ar) approach. DCL-Ar is innovated with a dual-layered protection mechanism that is adapted to varying architectures, protection requirements, and performance criteria concluded here in this work with three distinct operational scenarios. The methodology is capitalized on the collaborative potential of fog nodes to establish expansive cloak areas, effectively enhancing privacy without compromising service accuracy. Additionally, the strategic utilization of fog node cache has designed to mitigate potential performance detriments associated with protective measures. A comparative analysis, grounded in simulations, has demonstrated DCL-Ar's enhanced performance and protection levels over conventional methods such as the traditional Cloak-Area, the improved dummy, the obfuscation, and the developed collaboration approaches. The metrics for this comparative superiority encompass factors such as response time, accuracy, and the overall privacy-security balance.

However, the study acknowledges the inherent limitations within the DCL-Ar framework, including its reliance on the density and distribution of fog nodes, which may present challenges in sparse geographic areas. Furthermore, while the approach contributes a novel perspective to the discourse on data reliability in Crowdsourcing, a comprehensive exploration of this solution remains a topic for future research. Finally, the scalability of DCL-Ar and its practical application in real-world contexts are promising yet require further investigation to ascertain its viability across different scales and use cases. Security implications, particularly the resilience of DCL-Ar to various cyber threats, have not been the focus of this research and warrant additional scrutiny.

ACKNOWLEDGMENT

We thank the owners of Madinah date farms for their cooperation in enabling us to collect the dataset presented in this article.

REFERENCES

[1] F. Xia, L. T. Yang, L. Wang, and A. Vinel, "Internet of Things," *Int. J. Commun. Syst.*, vol. 25, no. 9, p. 1101, 2012.
[2] K. Pal, "Challenges of using wireless sensor network-based RFID technology for industrial IoT applications," in *Handbook of Research on Advancements of Contactless Technology and Service Innovation in Library and Information Science*. Hershey, PA, USA: IGI Global, 2023, pp. 80–100.
[3] H. M. A. Fahmy, "WSNs applications," in *Concepts, Applications, Experimentation and Analysis of Wireless Sensor Networks*. Cham, Switzerland: Springer, 2023, pp. 67–242.
[4] X. Jia, Q. Feng, T. Fan, and Q. Lei, "RFID technology and its applications in Internet of Things (IoT)," in *Proc. 2nd Int. Conf. Consum. Electron., Commun. Netw. (CECNet)*, Apr. 2012, pp. 1282–1285.
[5] Y. Hajjaji, W. Boulila, I. R. Farah, I. Romdhani, and A. Hussain, "Big data and IoT-based applications in smart environments: A systematic review," *Comput. Sci. Rev.*, vol. 39, Feb. 2021, Art. no. 100318.
[6] H. Cai, B. Xu, L. Jiang, and A. V. Vasilakos, "IoT-based big data storage systems in cloud computing: Perspectives and challenges," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 75–87, Feb. 2017.
[7] M. Marjani, F. Nasaruddin, A. Gani, A. Karim, I. A. T. Hashem, A. Siddiqa, and I. Yaqoob, "Big IoT data analytics: Architecture, opportunities, and open research challenges," *IEEE Access*, vol. 5, pp. 5247–5261, 2017.
[8] N. M. Bahbouh, S. S. Compte, J. V. Valdes, and A. A. A. Sen, "An empirical investigation into the altering health perspectives in the Internet of Health Things," *Int. J. Inf. Technol.*, vol. 15, no. 1, pp. 67–77, Jan. 2023.
[9] Y. Alsaawy, A. Alkhodre, A. Abi Sen, A. Alshanqiti, W. A. Bhat, and N. M. Bahbouh, "A comprehensive and effective framework for traffic congestion problem based on the integration of IoT and data analytics," *Appl. Sci.*, vol. 12, no. 4, p. 2043, Feb. 2022.
[10] F. H. Aljohani, A. A. Abi Sen, M. S. Ramazan, B. Alzahrani, and N. M. Bahbouh, "A smart framework for managing natural disasters based on the IoT and ML," *Appl. Sci.*, vol. 13, no. 6, p. 3888, Mar. 2023.
[11] H. Atlam, R. Walters, and G. Wills, "Fog computing and the Internet of Things: A review," *Big Data Cogn. Comput.*, vol. 2, no. 2, p. 10, Apr. 2018.
[12] A. A. A. Sen and M. Yamin, "Advantages of using fog in IoT applications," *Int. J. Inf. Technol.*, vol. 13, no. 3, pp. 829–837, Jun. 2021.
[13] N. M. Kumar and P. K. Mallick, "The Internet of Things: Insights into the building blocks, component interactions, and architecture layers," *Procedia Comput. Sci.*, vol. 132, pp. 109–117, Jan. 2018.
[14] H. Garcia-Molina, M. Joglekar, A. Marcus, A. Parameswaran, and V. Verroios, "Challenges in data crowdsourcing," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 4, pp. 901–911, Apr. 2016.
[15] G. Pasolini, A. Guerra, F. Guidi, N. Decarli, and D. Dardari, "Crowd-based cognitive perception of the physical world: Towards the Internet of senses," *Sensors*, vol. 20, no. 9, p. 2437, Apr. 2020.
[16] S. K. U. Zaman, A. I. Jehangiri, T. Maqsood, Z. Ahmad, A. I. Umar, J. Shuja, E. Alanazi, and W. Alasmary, "Mobility-aware computational offloading in mobile edge networks: A survey," *Cluster Comput.*, vol. 24, no. 4, pp. 2735–2756, Dec. 2021.
[17] W. Feng, Z. Yan, H. Zhang, K. Zeng, Y. Xiao, and Y. T. Hou, "A survey on security, privacy, and trust in mobile crowdsourcing," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2971–2992, Aug. 2018.
[18] S. Sodagari, "Trends for mobile IoT crowdsourcing privacy and security in the big data era," *IEEE Trans. Technol. Soc.*, vol. 3, no. 3, pp. 199–225, Sep. 2022.
[19] H. Li, L. Yu, and W. He, "The impact of GDPR on global technology development," *J. Global Inf. Technol. Manage.*, vol. 22, no. 1, pp. 1–6, Jan. 2019.
[20] H. Huang and G. Gartner, "Current trends and challenges in location-based services," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 6, p. 199, May 2018.
[21] M. Usman, M. R. Asghar, I. S. Ansari, F. Granelli, and K. A. Qaraqe, "Technologies and solutions for location-based services in smart cities: Past, present, and future," *IEEE Access*, vol. 6, pp. 22240–22248, 2018.
[22] A. A. A. Sen and A. M. Basahel, "A comparative study between security and privacy," in *Proc. 6th Int. Conf. Comput. for Sustain. Global Develop. (INDIACom)*, Mar. 2019, pp. 1282–1286.
[23] K. Y. Chai and M. F. Zolkipli, "Review on confidentiality, integrity and availability in information security," *J. ICT Educ.*, vol. 8, no. 2, pp. 34–42, Jul. 2021.
[24] A. A. Abi Sen, F. A. Eassa, K. Jambi, and M. Yamin, "Preserving privacy in Internet of Things: A survey," *Int. J. Inf. Technol.*, vol. 10, pp. 189–200, Jun. 2018.
[25] R. Gupta and U. P. Rao, "A hybrid location privacy solution for mobile LBS," *Mobile Inf. Syst.*, vol. 2017, pp. 1–11, 2017, doi: 10.1155/2017/2189646.
[26] C. Stamatellis, P. Papadopoulos, N. Pitropakis, S. Katsikas, and W. Buchanan, "A privacy-preserving healthcare framework using hyperledger fabric," *Sensors*, vol. 20, no. 22, p. 6587, Nov. 2020.
[27] W. Ren, X. Tong, J. Du, N. Wang, S. C. Li, G. Min, Z. Zhao, and A. K. Bashir, "Privacy-preserving using homomorphic encryption in mobile IoT systems," *Comput. Commun.*, vol. 165, pp. 105–111, Jan. 2021.
[28] P. Yang, N. Xiong, and J. Ren, "Data security and privacy protection for cloud storage: A survey," *IEEE Access*, vol. 8, pp. 131723–131740, 2020.
[29] I. Kreso, A. Kapo, and L. Turulja, "Data mining privacy preserving: Research agenda," *WIREs Data Mining Knowl. Discovery*, vol. 11, no. 1, p. e1392, Jan. 2021.
[30] H. Jin, Y. Luo, P. Li, and J. Mathew, "A review of secure and privacy-preserving medical data sharing," *IEEE Access*, vol. 7, pp. 61656–61669, 2019.
[31] 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.

[32] H. Xu, Y. Zhou, J. Ming, and M. Lyu, "Layered obfuscation: A taxonomy of software obfuscation techniques for layered security," *Cybersecurity*, vol. 3, no. 1, pp. 1–18, Dec. 2020.
[33] A. Diyanat, A. Khonsari, and S. P. Shariatpanahi, "A dummy-based approach for preserving source rate privacy," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 6, pp. 1321–1332, Jun. 2016.
[34] S. Siddiqie, A. Mondal, and P. K. Reddy, "An improved dummy generation approach for infeasible regions," *Int. J. Speech Technol.*, vol. 53, no. 15, pp. 18700–18714, Aug. 2023.
[35] 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, May 2014.
[36] J. Zheng, X. Tan, C. Zou, Y. Niu, and J. Zhu, "A cloaking-based approach to protect location privacy in location-based services," in *Proc. 33rd Chin. Control Conf.*, Jul. 2014, pp. 5459–5464.
[37] E. Fung, G. Kellaris, and D. Papadias, "Combining differential privacy and PIR for efficient strong location privacy," in *Proc. Int. Symp. Spatial Temporal Databases*. Cham, Switzerland: Springer, Aug. 2015, pp. 295–312.
[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] F. Kreuter, G.-C. Haas, F. Keusch, S. Bähr, and M. Trappmann, "Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent," *Social Sci. Comput. Rev.*, vol. 38, no. 5, pp. 533–549, Oct. 2020.
[40] A. Frik, J. Kim, J. R. Sanchez, and J. Ma, "Users' expectations about and use of smartphone privacy and security settings," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Apr. 2022, pp. 1–24.
[41] 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.
[42] 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.
[43] C. O. Adetunji, O. T. Olaniyan, O. Adeyomoye, A. Dare, M. J. Adeniyi, E. Alex, and M. A. Shariati, "Internet of Health Things (IoHT) for COVID-19," in *Assessing COVID-19 and Other Pandemics and Epidemics Using Computational Modelling and Data Analysis*. Cham, Switzerland: Springer, 2022, pp. 75–87.
[44] C. Zhang, L. Zhu, C. Xu, X. Du, and M. Guizani, "A privacy-preserving traffic monitoring scheme via vehicular crowdsourcing," *Sensors*, vol. 19, no. 6, p. 1274, Mar. 2019.
[45] M. da Silva, J. Viterbo, F. Bernardini, and C. Maciel, "Identifying privacy functional requirements for crowdsourcing applications in smart cities," in *Proc. IEEE Int. Conf. Intell. Secur. Informat. (ISI)*, Nov. 2018, pp. 106–111.
[46] B. Rashidi, C. Fung, A. Nguyen, T. Vu, and E. Bertino, "Android user privacy preserving through crowdsourcing," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 3, pp. 773–787, Mar. 2018.
[47] N. H. Hassan and F. A. Rahim, "The rise of crowdsourcing using social media platforms: Security and privacy issues," *Pertanika J. Sci. Technol.*, vol. 25, no. 110, pp. 79–88, 2017.
[48] M. Yamin and A. A. Abi Sen, "A new method with swapping of peers and fogs to protect user privacy in IoT applications," *IEEE Access*, vol. 8, pp. 210206–210224, 2020.

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 pursuing the Ph.D. degree in the Internet of Health Things (IoHT) domain with Grandad University, Spain. She has published more than 20 researches most of them in the employing IoT and its applications to service society. In addition, she has good experience in design frameworks and simulation.

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 Madinah. His research interests include software engineering, intelligent systems and cities, sensing networks privacy, security, and embedded real-time systems.

SANDRA SENDRA (Member, IEEE) is currently a Professor with the Communication Department, Universitat Politècnica de València, Spain. She has more than 160 publications in different domains and many of them are published in high-quality journals (ISI).

ADNAN AHMED ABI SEN received the Ph.D. degree in computer sciences from King Abdul-Aziz University (KAU), Saudi Arabia. In his academic career, he has supervised dozens of graduation projects in the various fields of computer science. He is an experienced and established researcher in the IoT applications, privacy, and mobile computing. He also has expertise in systems analyzes and development. He is a Researcher and Software Manager at the Digital Transformation and Smart City Agency, Madinah Regional Municipality, Saudi Arabia, for the Perfect Presentation Company (2P). He has published about 70 research articles, many of which are indexed in the Thomson and Reuters databases. He received the Best Thesis Award for the Ph.D. degree.

YAZED ALSAAWY received the Ph.D. degree in computer science from De Montfort University, U.K., in 2014. He was the General Manager of the Digital Transformation Department, Ministry of Education, and a Leader of a package of projects and initiatives at different levels in the Ministry of Education. Before that, he was the Dean of the Information Technology Deanship, Islamic University of Madinah. He works as an Assistant Professor with the Faculty of Computer and Information Systems, Islamic University of Madinah, and spent more than a year as the Vice Dean for Academic Affairs at the Faculty of Computer and Information Systems. He is currently an Associate Professor. He participated in Stanford Executive Program (SEP), in 2018. He works in the fields of security, privacy, and the IoT. He is the co-author of many articles in software engineering, eLearning, networking, blockchain, security, and privacy. He participated in many founded projects. He submitted many research projects for funding at King Abdulaziz City for Science and Technology (KACST).

MOHAMED BENAIDA received the Ph.D. degree in computer science. He is currently an Associate Professor with the Islamic University of Madinah. His research interests include HCI usability web design e-government and computing education.

HANI ALMOAMARI received the Ph.D. degree in computer science from the University of Stirling. He is currently an Assistant Professor with the Faculty of Computer and Information Systems, Islamic University of Madinah, and the Vice Dean for Quality Assurance. He is the co-author of many articles in software engineering, eLearning, networking, blockchain, security, and privacy. His current research interests include network security, privacy, and the IoT.