Preserving Location Privacy in the IoT against Advanced Attacks using Deep Learning

Abdullah S. Alyousef[1]*, Karthik Srinivasan[2], Mohamad Shady Alrahhal[3], Majdah Alshammari[4], Mousa Al-Akhras[5]
College of Computing and Informatics, Saudi Electronic University, Riyadh, Saudi Arabia[1, 2, 5]
Department of Computer Science, King Abdulaziz University, Jeddah City, Saudi Arabia[3]
Department of Computer Science, Hail University, Hail City, Saudi Arabia[4]
Computer Information Systems Department, King Abdullah II School for Information Technology[5]
The University of Jordan, Amman, Jordan[5]

***Abstract*—Location-based services (LBSs) have received a significant amount of recent attention from the research community due to their valuable benefits in various aspects of society. In addition, the dependency on LBS in the performance of daily tasks has increased dramatically, especially after the spread of the COVID-19 pandemic. LBS users use their real location to build LBS queries to take benefits. This makes location privacy vulnerable to attacks. The privacy issue is accentuated if the attacker is an LBS provider since all information about users is accessible. Moreover, the attacker can apply advanced attacks, such as map matching and semantic location attacks. In response to these issues, this work employs artificial intelligence to build a robust defense against advanced location privacy attacks. The key idea behind protecting the location privacy of LBS users is to generate smart dummy locations. Smart dummy locations are false locations with the same query probability as the real location, but they are far from both the real location and each other. Relying on the previous two conditions, the deep-learning-based intelligent finder ensures a high level of location privacy protection against advanced attacks. The attacker cannot recognize the dummies from the real location and cannot isolate the real location by a filtering process. In terms of entropy (the privacy protection metric), accuracy (the deep learning metric), and total execution time (the performance metric) and compared to the well-known DDA and BDA systems, the proposed system shows better results, where entropy = 15.9, accuracy = 9.9, and total execution time = 17 sec.**

***Keywords*—LBS; dummy; deep-learning; attacks; accuracy; resistance; performance**

I. Introduction

The Internet of Things (IoT) can be defined as a network of devices that are connected through the Internet to facilitate performing tasks remotely. The IoT is involved in all aspects of people's lives, and it can be used in a wide range of applications in industry, transportation, and medicine [1]. In smart cities, the IoT forms the backbone for performing several missions, as shown in Fig. 1.

Among the IoTs, location-based services (LBSs) are considered the most important services that serve people daily. LBSs can be seen as commercial location applications that utilize the geographical location information of smart devices and mainly smartphones, enabling users to search for Points of Interest (PoIs), such as nearest restaurants, hospitals, libraries, and sports clubs [3]. In other words, LBSs employ a Global Positioning System (GPS) to perform queries issued from the user side. In addition, smartphone users can easily obtain the benefits of LBS applications by downloading them from various sites, such as the Apple Store or Google Play Store. From an intersection of technologies point of view, LBS can be illustrated as shown in Fig. 2.

A. The Importance of Location-based Services

In general, the importance of LBSs comes from their provided benefits, which make our lives easier and more enjoyable. In detail, three main sectors of daily life highlight the importance of LBS-enabled applications:

1) Medical sector. In the e-health field, LBSs play a significant role in monitoring patient health conditions (e.g. pulse rates and blood pressure levels), avoiding disasters [4, 5]. This, in turn, means that LBSs contribute to limiting the spread of illnesses such as COVID-19 by enabling medical staff and patients to avoid meeting and consequently maintaining a safe social distance.

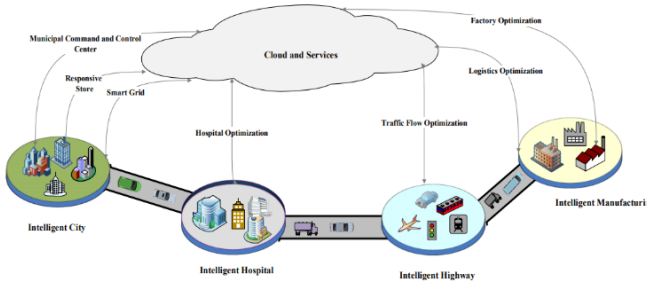

Fig. 1. IoT in Smart Cities [2].

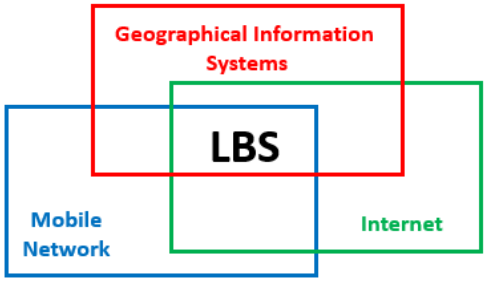

Fig. 2. LBS from an Intersection of Technologies Point of View.

*Corresponding Author.

2) Entertainment sector. A further advantage of LBSs is enabling users to search for PIOs, such as nearby restaurants and music clubs, or enjoying games online [6, 7].

3) Social sector. Integrating LBS applications with wireless communication technologies have enabled the creation of location-based social networking services, such as Foursquare, Twinkle, and GeoLife [8]. This integration bridges the gap between the physical world and digital online social networking services.

B. Statement of Problem

The valuable benefits of the LBS applications mentioned above are not without risk. The key problem behind the extensive utilization of LBS applications is that the privacy of LBS users may be attacked. In the cybersecurity research field, privacy is a term that refers to sensitive information about users' interests, habits, or personal lives [9, 10]. Obtaining such information harms users and can even threaten their lives in cases of blackmailing or stealing valuable personal information, including the nature of their business, the details of their business trips, or their religious affiliation.

To gain a deep look at the privacy issue in LBS applications, the mechanism used for serving users should be analyzed. Using LBS applications requires constructing and sending queries relying on the real geographical locations of LBS users, who obtain their real locations through GPS. After manipulating these queries by the LBS provider, the results are returned to the users. Fig. 3 illustrates the general mechanism followed by LBS applications.

As shown in Fig. 3, there are three main steps, as follows.

1) The LBS user establishes a query using their real location. This query is then sent to the LBS provider.

2) The LBS provider processes the received query to answer the user. The result of the query (the retrieved POI) is packaged for resending.

3) The result is sent back to the LBS user and seen on the smartphone screen.

The scenario described in Fig. 3 is insecure against an attacker targeting the privacy of the LBS user. To define the problem accurately, modelling is required. Let $< \alpha.\beta >$ denote the coordinates of the real location of a given LBS user. Based on this representation, the query that is sent to the LBS provider is defined as:

$$Q_{LBS} = \{< \alpha.\beta >.S_{PoI}.D.ID\} \quad (1)$$

Where: S_{PoI}: set of points of interest that represent the result of the sent query. D: diameter of the search region (measured by Kilometres). ID: identity of the LBS user.

The privacy problem starts when an attacker tracks the real location of the LBS user or analyzes the sent query, as shown in Fig. 4.

In both cases (i.e., tracking the real location or analyzing the sent query), personal information about the LBS user is obtained. Malicious activities can be performed by a man-in-the-middle (MITM) attack. However, the privacy problem is accentuated if the attacker is the LBS provider since all information is accessible. Upon this, the attacker (the LBS provider) can track the real location of the LBS user or analyze the received query. A malicious profile is then constructed on the attacker side, containing personal information that will be employed to attack the victim physically. Fig. 5 illustrates this dangerous scenario.

Formally, let VP denote the victim profile. Then,

$$VP = Track\ (< \alpha.\beta >) \cup (Analyze\ Q_{LBS}) \quad (2)$$

In terms of data flow and trust boundaries (attack surface), the security gap is represented by obtaining the query illustrated in Fig. 6.

It is worth mentioning that tracking the real location of LBS users leads to location privacy issues, and analyzing the sent query leads to query privacy issues [11]. In this work, we are concerned about location privacy only.

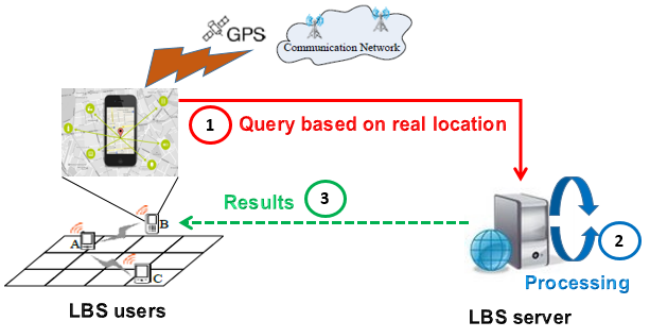

Fig. 3. The General Mechanism followed by LBS Applications.

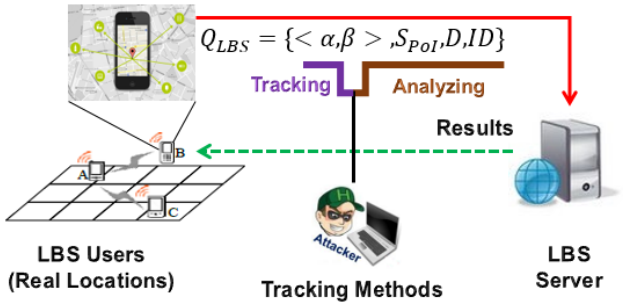

Fig. 4. Privacy Problem in LBS Applications.

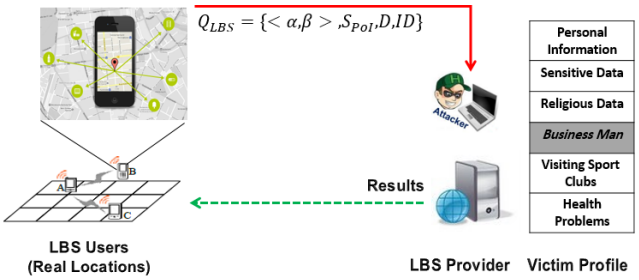

Fig. 5. Accentuated Privacy Problem in LBS Applications (LBS Provider is Attacker).

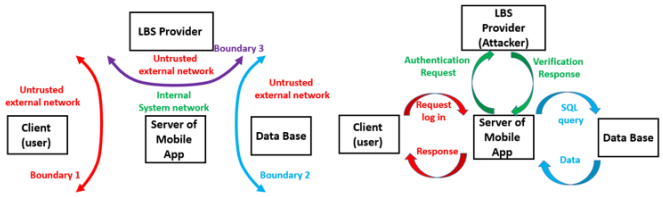

Fig. 6. Security Gap from the Attack Surface Perspective.

C. Motivation and Research Questions

In light of the dangerous spread of the COVID-19 pandemic, dependence on mobile applications and the Internet has increased. This is because this dependency keeps people healthy in terms of achieving social distancing requirements. Increasing dependency on mobile applications is tightly coupled with an increasing level of privacy threats [12]. Moreover, existing advanced methods that could be used to track users, such as those that gather private information [13, 14], make privacy concerns more relevant. The capabilities of attackers are growing daily, with advanced attacks used to collect personal information from LBS users being applied. The attacker in a Map Matching Attack (MMA) employs the side information to gather sensitive data about the LBS user. In other words, the attacker can discover the kinds of activities the user is involved in by knowing the geographical map from which the LBS query is issued (i.e., without tracking the real location of the LBS user) [15, 16]. Fig. 7 illustrates the basic concept of an MMA.

Another advanced attack used for penetrating protection methods is the Semantic Location Attack (SLA) [18]. In an SLA, the attacker can infer semantic meanings related to the user's behavior, relying on both the time and place of where a user stays, as shown in Fig. 8.

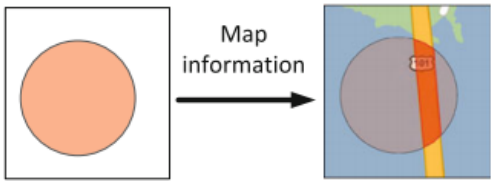

Fig. 7. Concept of MMA [17].

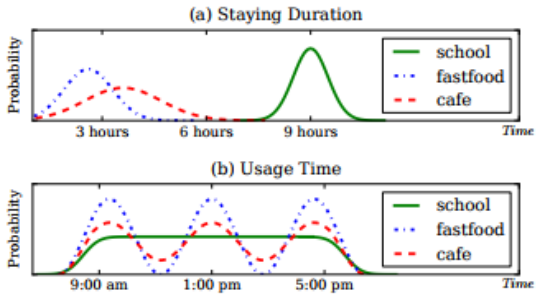

Fig. 8. Concept of SLA [19].

Motivated by these advanced attacks, two main research questions must be answered:

1) How do we ensure high resistance against both MMA and SLA?

2) How can privacy protection be quantified in terms of preventing attackers from penetrating privacy protection approaches?

D. Contribution

The contributions of this work are listed as follows:

- In response to the first research question, a deep learning technique is proposed to generate strong dummy locations that protect the real location of the LBS user.
- In response to the second research question (second quality requirement), the entropy metric is employed to measure the resistance of the proposed deep learning based privacy protection system.

E. Structure of the Work

The rest of this work is organized as follows. Related work is reviewed in Section II. Section III presents the methodology of designing and constructing the proposed system in detail. Security analysis is discussed in Section IV, followed by the results in Section V. Finally, the conclusion and suggestions for future work are provided in Section VI.

II. Related Work

In response to privacy concerns, researchers have proposed several approaches. The approaches were addressed from different perspectives, namely, server-based approaches, user-based approaches, and Trusted Third Party (TTP) approaches. Fig. 9 is a classification of LBS privacy protection approaches, where each category has its drawbacks.

The authors of work [20] proposed a Dummy Data Array (DDA) algorithm for generating dummy locations to protect the location privacy of LBS users. For a given region, which is divided into a grid of cells, the key idea of the DDA algorithm is to calculate both the vertices and the edges of each cell in the grid. Then, the DDA algorithm randomly selects some of the cells as dummy locations. To select strong dummy locations and achieve k-anonymity, the DDA algorithm selects k cells of equal area. The authors of the work [22] provided a survey of privacy protection approaches and they focused on dummies. Similarly, [21] uses dummies to protect the location privacy of LBS users, but with a different dummy generation method. The authors proposed two algorithms. The first is called CirDummy, which generates dummies based on a virtual circle that contains the real location of the LBS user. The second is called GridDummy, which generates dummies based on a virtual grid that covers the real location of the LBS user.

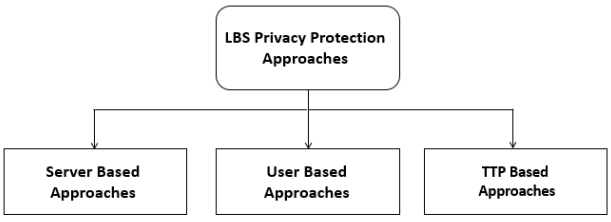

Fig. 9. Classification of Privacy Protection Approaches for LBS Applications.

Mix zones are defined in which all users' locations are hidden within these zones with some conditions to strengthen the protection method. In the work [23], the authors present the (DSC-LPP) approach to protect location privacy in the wireless channels. It is based on the idea of spatial cloaking-location privacy preserving. In the wireless channels, access points are essential elements in the structure of the network. Distance between the user and AP is the primary key for transforming and retrieving the location of users. Relying on this fact, the authors proposed to include the distance information in exchanged messages. The distance information can be exploited to confuse the attacker by manipulating it through mathematical transforming. The transformation leads to calculate a new location information, which in turn forms a clocking region. The clocking region reflects a security area of the location privacy. To enhance the performance, normalization is employed in the process of transformation for the purpose of building clocking regions. The advantage of this approach is that it can protect location privacy in a fixable way depending on increasing or decreasing the area of the clocking regions. However, if attackers have side information about the geographical map where the LBS user is located, the mechanism of protection becomes weak. In other words, this approach is not robust against MMA.

Pseudonym method [24] is used for protection of the user identity. The key idea is confusing the relationship between the position information and user identity information. This method is based on TTP model, TTP is the simplest intermediary entity between the user and the LBS provider. If the request is accepted, the request will be sent to the LBS provider; at the same time, the real ID will be changed to a pseudo-ID.

Information Retrieval (PIR) [25] was used to achieve full privacy protection. The key idea of the PIR technique depends on mathematical principle. It says that if it is impossible to compute a certain number or perform a certain mathematical task, then the information that form the task is protected. When the query is represented by a task, and PIR technique is applied, then, the LBS server can process and answer the query without knowing any sensitive information about the query.

III. Proposed System

This section is structured so that the threat model is defined first. The proposed system design is then described in detail. Next, security analyzes are discussed. Finally, the mechanism of evaluation of the proposed system is presented with the corresponding metrics.

A. Threat Model

The objective of the threat model is to draw the environment within which the proposed system is running and is expected to be robust against attackers. The threat model consists of four blocks as shown in Fig. 10.

- Attacker. The attacker is the LBS provider itself (or its maintainer), where all LBS queries are sent to it, and connecting with this malicious party is mandatory.
- Malicious goal. The goal of the attacker is to build a malicious profile about the LBS user. This is done by gathering personal data about the victim through tracking the real locations used to establish LBS queries.
- Capabilities. The attacker's ability is supported by launching attacks on the victim, including MMA and SLA attacks.
- Type of attack. The type of each attack launched on the victim is active. This is because the LBS provider (attacker) can access all information received while serving the LBS user.

B. System Design

This section provides the architecture of the proposed system with its main components and the role of each component.

1) Architecture of the proposed system: The system decomposes three main components: the intelligent finder, query builder, and sender. The system is decentralized one because it is installed on each mobile device of LBS user. Table I summarizes the three components in terms of the assigned task, technique used, and installation.

Graphically, Fig. 11 shows the architecture of the proposed system with interconnections among the three components.

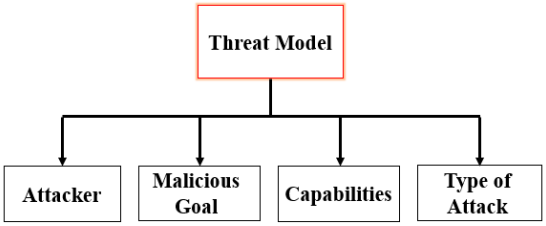

Fig. 10. Blocks of the Threat Model.

TABLE I. Components of the System

Name	Task	Technique	Installation
Intelligent finder	Generating dummy locations	Convolutional Neural Network (CNN), Support Vector Machine (SVM)	LBS user (Smartphone)
Query builder	Building the protected query	Anonymity of identity	LBS user (Smartphone)
Sender	Sending the protected query	Wireless communication	LBS user (Smartphone)

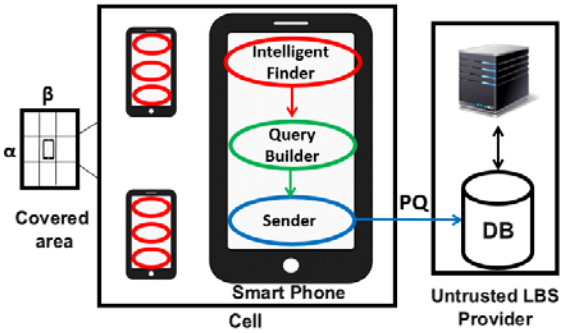

Fig. 11. Architecture of the Proposed System.

As shown in Fig. 11, the smartphone (which represents an LBS user) is located in a certain location (cell) within the covered area that consists of (α×β) cells. The rest of the cells are spread on different regions that contain various PoIs. The cells that form the covered area can be exploited as dummy locations to protect the location privacy of LBS users. In other words, using fake locations instead of the real location cuts the tracking series that is performed on the attacker's side to complete the malicious profile. This is because the attacker (LBS provider) cannot recognize the real location among dummies. However, the attacker attempts to compromise the protection method by applying advanced attacks such as MMA and SLA. This requires that the process of generating (or finding) dummy locations be accurate to provide strong dummies that can protect location privacy against advanced attacks. In this work, artificial intelligence is employed to generate strong dummies based on Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs). Upon this, the intelligent finder selects (generates, or searches) strong dummy locations and then provides them to the query builder to establish a query with multiple locations (one of them is the real one). Then, the sender sends the protected query (PQ) to the LBS provider (attacker). The attacker is confused about determining the real location among dummies. Below is a detailed description of the role of each component.

2) Role of an intelligent finder: The main task of this component is ensuring the location privacy of the LBS user. This is performed by protecting the location information used to form the sent query. Based on a novel location privacy protection approach, namely, Vectors of Protection (VoP), this component ends its assigned task. VoP fills a vector of locations by dummies, and the real location in the LBS query is replaced by this vector. The key idea of the VoP approach is illustrated in Fig. 12.

As shown in Fig. 12, the real location of the LBS user (LBS_L^r) is represented by the left side. The role of the intelligent finder component is to fill the vector by dummy locations by executing the VoP approach. The rest of the query units remain constant.

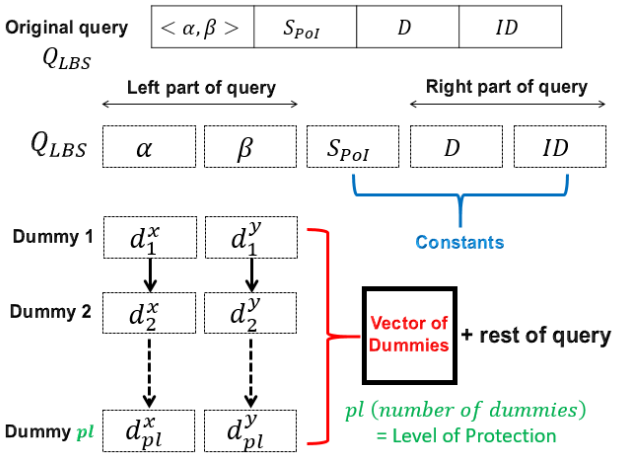

Fig. 12. Key idea of the VoP Approach.

In detail, for a region divided into $\alpha \times \beta$ cells, the real location of the LBS user LBS_L^r is located in a certain cell. Each cell has a query probability $CELL_p^q$. The query probability is a term that refers to the number of queries sent from a specific location in the past (i.e., number of queries built based on the cell divided by the total number of queries built based on the whole cells). Each cell has a certain value of query probability, as shown in Fig. 13.

The VoP approach selects dummies randomly. From the real location of the LBS user, some vectors are issued to the selected dummies. Then, the selected dummies are stored in the vector of dummies. The number of dummies determines the level of protection. This means that the LBS user has full control over the desired level of privacy protection. For instance, if the LBS user selects 3 dummy locations, the level of privacy protection is 4. This is because the real location is surrounded by three dummies, as shown in Fig. 14.

The process of selecting dummies randomly without any constraint is a poor tactic. This is because the query probability of each dummy location differs from the query probability of the real location of the LBS user. This increases the ability of the attacker to determine the real location among dummies. Therefore, it is better to select dummy locations with the same query probabilities as the real location of the LBS user. Fig. 15 illustrates the selection process under the same query probability condition.

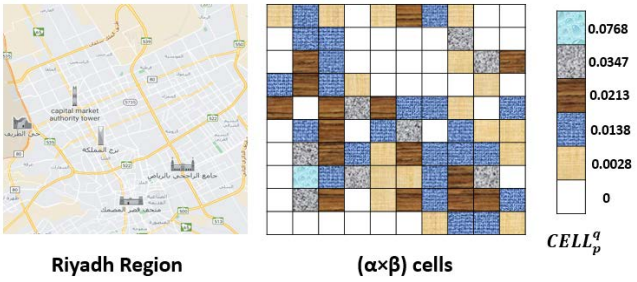

Fig. 13. Query Probabilities of Cells.

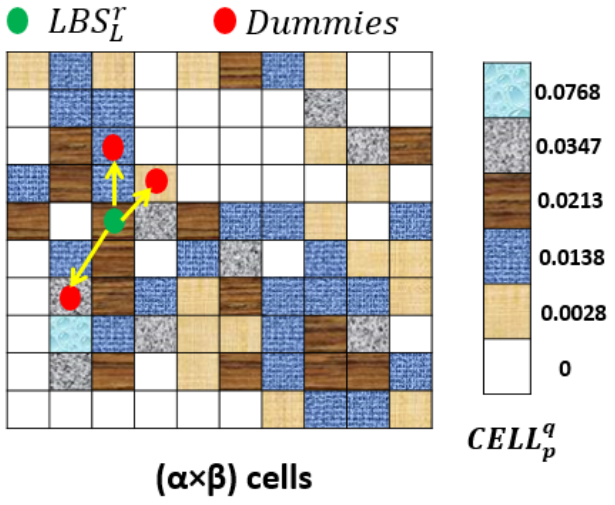

Fig. 14. Achieving Privacy Protection of 4 Levels.

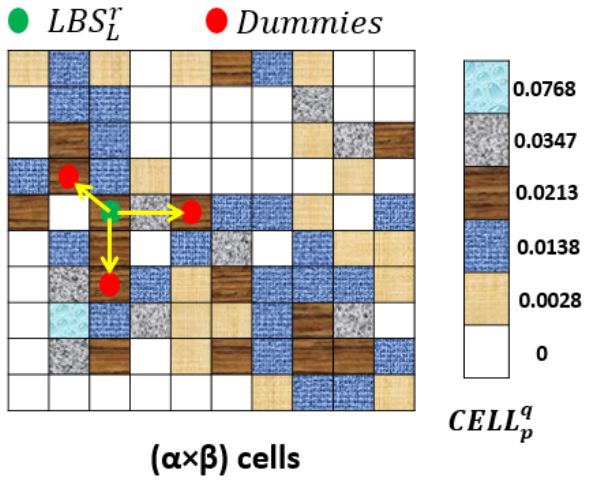

Fig. 15. Selecting Dummies Depending on the Same Query Probability Condition.

This method guarantees that the uncertainty (at the attacker's side) in determining the real location among the dummies is maximal. Mathematically, this uncertainty is represented by entropy. Entropy is a term that refers to the inability to determine an object among others based on the same features [26]. The entropy of identifying the real location out of the dummy vector ENT_{dv}^r is defined as:

$$ENT_{dv}^r = -\sum_{i=1}^{pl} CELL_{p_i}^q \times \log_2 \times CELL_{p_i}^q \quad (3)$$

where pl denotes the protection level of privacy.

Despite selecting dummies based on the query probabilities condition, the privacy threat remains. Selecting weak dummy locations creates a vulnerability where the attacker can apply MMA and SLA successfully. Weak dummies mean that the dummy locations are near the real location of the LBS user. This allows the attacker to filter dummies easily. This requires additional conditions in the process of selecting dummy locations. This condition states that the selected dummies must be far away from the real location of the LBS user, as shown in Fig. 16.

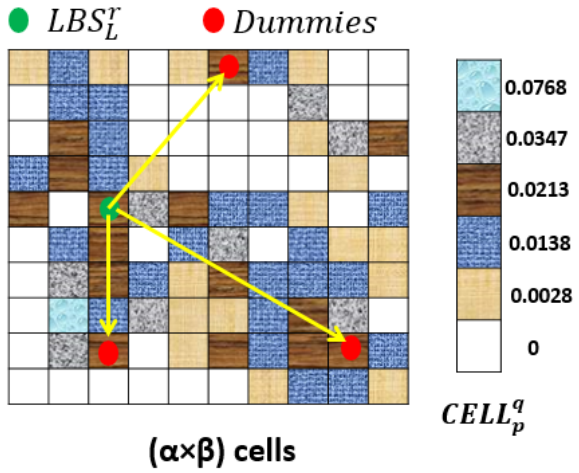

Fig. 16. Selecting dummies Depending on both the same Query Probabilities and Far away Conditions.

The actual selection process is performed by the intelligent finder component. Electing suitable dummies (i.e., strong dummies) requires an intelligent method. This intelligent method depends on scanning the covered region and then determining strong dummies. In this work, a deep learning method (the CNN network) with the help of SVM is employed to elect strong dummies to fill the vector of dummies.

The task of the CNN is extracting the features of a given geographic region (map). This is performed by scanning the map through two kinds of layers: convolutional layers and pooling layers. Fig. 17 illustrates the mechanism used by the CNN for extracting features of a given region.

As shown in Fig. 17, the CNN goes through the map in a convolutional manner (depending on a filter or kernel) to extract the first level of features. Then, the extracted features are grouped through a pooling layer to draw a deep look at the locations included in the map. This procedure (i.e., convolution and pooling) is repeated frequently for the series of extractions. The final pooling layer includes the final features. Among the extracted features, some locations are suitable to be strong dummies, while some are weak dummies. The SVM is linked to the fully connected layer to classify the locations into two main groups: strong dummies and weak dummies. SVM is an intelligent technique that separates a given set of data into two major classes. SVM relies on margin, which can be seen as a restricted area between the two classes. Fig. 18 shows the basic concept of SVM.

SVM is represented mathematically by the sigmoid function, which forms the (S) curve from a graphical perspective, as shown in Fig. 19.

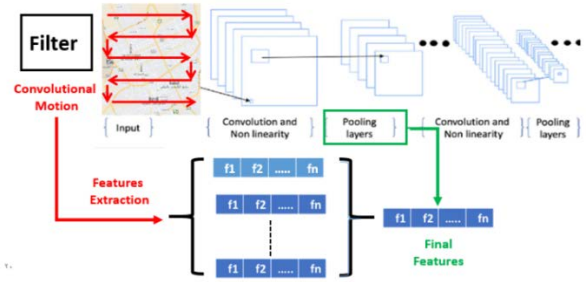

Fig. 17. Extracting Features of Map using CNN.

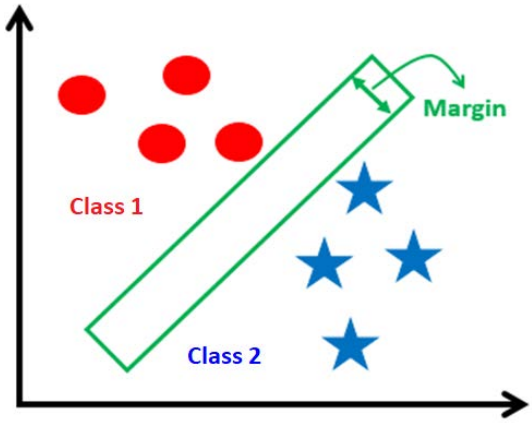

Fig. 18. Basic Concept of SVM.

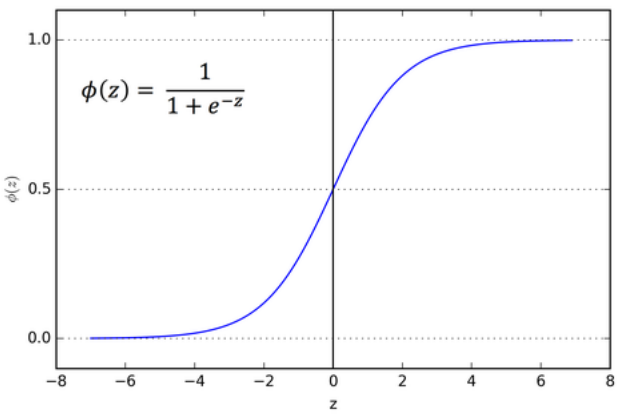

Fig. 19. Sigmoid Function.

According to the sigmoid function, the complete CNN network will be seen as a classifier, where all values above (+1) or lower than (0) represent strong dummies, and the area between the range [0, +1] represents the margin. Fig. 20 shows the complete CNN.

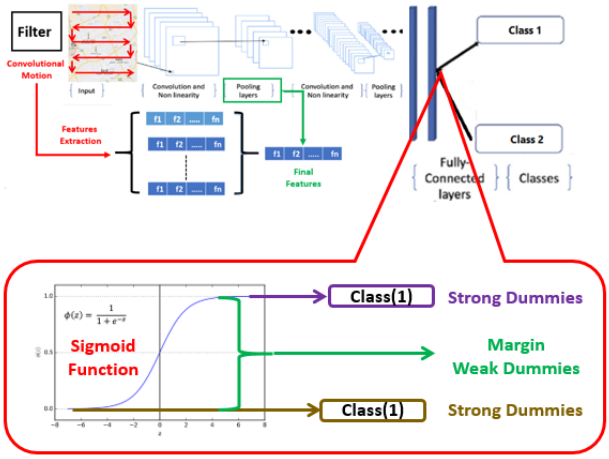

Fig. 20. Complete CNN Network.

Regarding the VoP approach, the complete CNN feeds it with a set of strong dummies. Let $final_{CNN}^{dummies}$ denotes the final set of strong dummies. The vector of dummies is then filled by a random selection of dummies since they all satisfy the two conditions. The size of the VoP (or the number of selected dummies) is based on the privacy protection level that is desired. Mathematically,

$$VoP = random\{final_{CNN}^{dummies}\} \quad (4)$$

The intelligent finder represented by the CNN is trained on the Brightkite dataset [27]. It consists of 7.3 million rows and five columns (user ID, check-in time, latitude, longitude, and location id). To involve query probabilities, we add a new column QP to the database. The values of the QP are generated randomly. Additionally, the Brightkite dataset is used for the testing stage. Therefore, the dataset is divided into two parts, as shown in Fig. 21.

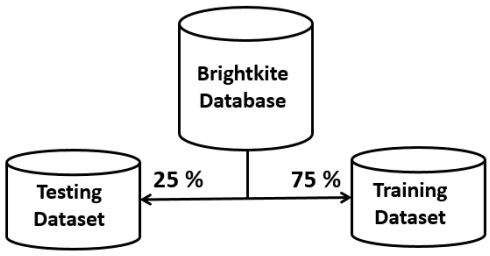

Fig. 21. Dividing the Brightkite Dataset.

3) Role of the query builder: This component is responsible for building the protected query. It receives the vector of dummies generated by executing the VoP approach (the task of the intelligent finder) and then constructs the query. To add a second layer of privacy protection, another task is assigned to this component, which blurs the ID of the LBS user. To end this, the query builder components use an anonymity technique. The key idea behind the anonymity technique is to hide the ID of the LBS user by replacing it with a fake ID. Upon this, the query generated by the query builder component is constructed as shown in Fig. 22.

As shown in Fig. 22, the ID of the LBS user in the original query is replaced by a fake identity ($\widetilde{ID}$). This adds additional protection to location privacy since the attacker (LBS provider) can recognize neither the real location among dummies nor the identity of the LBS user. Thus, the whole units of the protected query are given by:

$$Protected[Q_{LBS}] = \{< \begin{matrix} \alpha & \beta \\ d_1^x & d_1^y \\ d_2^x & d_2^y \\ \vdots & \vdots \\ d_{pl}^x & d_{pl}^y \end{matrix} >. S_{PoI}.D.\widetilde{ID}\} \quad (5)$$

4) Role of the sender: This component is responsible for sending the protected LBS query to the LBS provider for manipulation. Fig. 23 illustrates the task of the sender component.

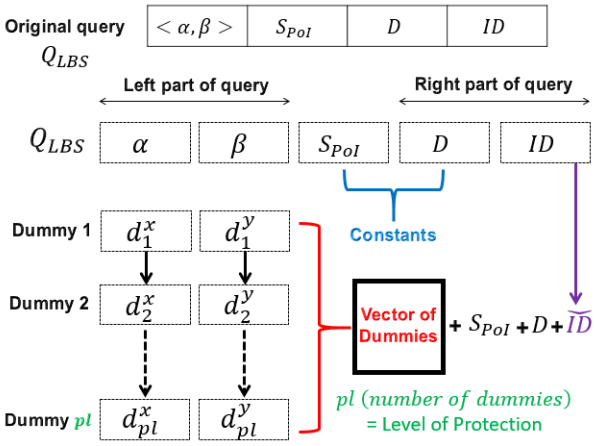

Fig. 22. The Query Builder Constructing a Protected Query.

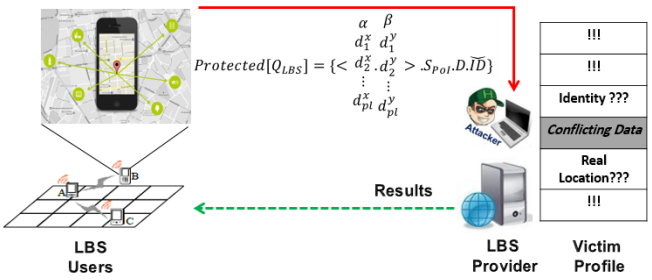

Fig. 23. Sending a Protected Query to the LBS Provider (Attacker).

Manipulating the protected LBS query at the LBS provider-side leads to confusion. This is because the victim profile will be full of conflicting data that is not beneficial. Thus, the location privacy of the LBS user is protected against the LBS provider if it acts maliciously.

IV. Security Analysis

This section discusses the security analyses, where the issues are taken from the attacker's perspective. In this context, two main issues are involved, as shown in Fig. 24.

In regard to knowing the proposed approach by the attacker, a reversing trial is expected to be performed to filter dummies. Here, the power of the randomization process is employed. As mentioned in formula 4 above, the final set of dummies is selected randomly. This means that if 20 dummies are strong and can be used as actual dummies, 5 dummies can be selected randomly to be utilized as actual dummies to achieve a protection level of 6 degrees. Randomization ensures complete doubt about determining which of the 5 dummies are elected among the 20 dummies. This reflects uncertainty in the process of selecting dummies on the attacker side. As a result, the attacker can only randomly guess the real location among dummies. Thus, reversing the VoP approach fails to achieve the malicious goal of the attacker.

In regard to discussing the success of the MMA and SLA attacks from the perspective of the attacker, these attacks fail. The reason is that the VoP approach takes into account the $CELL_p^q$, where it is the same for all locations (the real and dummies). In addition, the dummies selected by the CNN are far from both each other and the real location. Therefore, attempting to collect the dummies in one region is inapplicable at the attacker side. This means that the success of the attacks will not be achieved.

Fig. 24. Issues of Security Analysis.

V. Result and Discussion

This section provides the results in the context of comparison with two approaches. The first approach is the classical one. A classical approach is an approach inspired by the proposed VoP approach, where the dummies are selected randomly without taking any optimization into account (i.e., $CELL_p^q$ and distance-relation between the LBS_L^r and the other dummies). The classical approach is referred to as the basic dummy approach (BDA). The second approach involved in the comparison is the one that is proposed in ref [20], which is DDA.

A. Evaluation Metrics

In this work, three types of metrics are used for evaluation, as shown in Fig. 25.

The privacy-based metric, entropy, which is defined above (by formula 3), is employed to measure the privacy protection level that is achieved. Entropy is a metric addressed by many authors who conducted surveys, such as [28- 32], and by others who made technical research papers, such as [33-36]. The mechanism of an evaluation relying on the entropy metric is adjusted by the following rules:

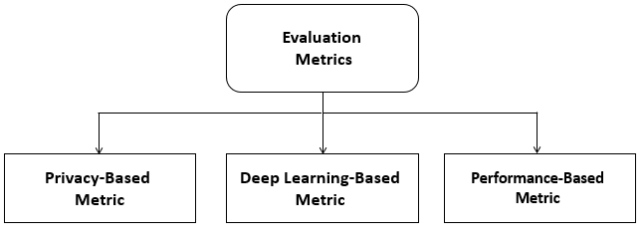

Fig. 25. Types of Evaluation Metrics.

1) There is no upper limit to the value of entropy.
2) There is no lower limit to the value of entropy.
3) A higher entropy value means a higher privacy protection degree.
4) A lower entropy value means a lower privacy protection degree.

For the deep-learning-based metric, an accuracy metric is utilized. Accuracy is a term that refers to the ratio of records (dummies) that are correctly classified (i.e., selected as strong dummies) [37]. The accuracy metric is inspired by a confusion matrix, a common term in the data mining research field [38]. Table II shows the confusion matrix (COfMX).

TABLE II. COfMX and its Components

Actual dummy (Predicted dummy)	**Confusion matrix**		
	DM	**¬ DM**	**Sum**
DM	True positives DM (TPDM)	False negatives DM (FNDM)	TPDM + FNDM = P
¬ DM	False positives DM (FPDM)	True negatives DM (TNDM)	FPDM + TNDM = N

Where:

1) TPDM is a positive dummy that is correctly labelled by the CNN classifier.

2) TNDM is a negative dummy that is correctly labelled by the CNN classifier.

3) FPDM is a negative dummy that is incorrectly labelled positive.

4) FNDM is a positive dummy that is mislabelled negative.

Accuracy is given by the following formula:

$$Accuracy = \frac{(TPDM+TNDM)}{number\ of\ all\ records/dummies\ in\ the\ testing\ set} \quad (6)$$

The mechanism of evaluation relying on the accuracy metric is adjusted by the following rules:

1) There is an upper limit to the value of accuracy (1 or 100%).

2) There is a lower limit to the value of accuracy (0).

3) A higher accuracy value means a higher prediction degree.

4) Lower accuracy value means a lower prediction degree.

For the performance-based metric, time dominates the case. Thus, the total execution time ($TexeT$) required to execute the approach is used. The $TexeT$ is defined by:

$$TexeT = T_{VoP}^{exe} + 2 \times T_{query}^{send} + T_{query}^{processing} \quad (7)$$

where T_{VoP}^{exe} refers to the time of executing the proposed VoP approach at the smartphone of the user, T_{query}^{send} refers to the sending time of the query (assuming that the return takes the same time), and $T_{query}^{processing}$ refers to the processing time at the server-side to answer the send query. The mechanism of evaluation relying on the $TexeT$ metric is adjusted by the following rules:

1) There is no upper limit to the value of $TexeT$.

2) There is no lower limit to the value of $TexeT$.

3) A higher $TexeT$ value means a lower performance degree.

4) A lower $TexeT$ value means a higher performance degree.

B. Entropy-Based Evaluation without Threats

Without applying any threat, the value of entropy is calculated to increase the protection level from 3 to 21 (i.e., from three dummies to 21 dummies). Fig. 26 shows the results.

Discussion and justifications: As shown in Fig. 26, the CNN-VoP and BDA approach experiences increased entropy values as the protection level increases. The reason is related to mathematical justification, where increasing the number of dummies involved in the protection level leads to an increase in the entropy value. However, the CNN-VoP approach achieves better scores than the BDA approach. This is due to selecting dummy locations under the control of $CELL_p^q$ in the CNN-VoP approach. In contrast, no constraints are used in the BDA approach. For the DDA approach, its curve can be divided into three parts. The first part behaves the same as the CNNN-VoP and BDA approaches to increase the values of entropy. In the first part, the DDA sometimes overcomes the BDA depending on the tree that combines similar dummies, which in turn means that some dummies have $CELL_p^q$ that are similar to the $CELL_p^q$ of the real location or approximately close to it. In the second part, where PL=12, the DDA performs the worst. This is because there are no available candidates that can be used as actual dummies, and in this case, the DDA repeats the dummies, which negatively affects the entropy value. In the third part, the DDA enhances slightly, but the BDA outperforms it due to the broad set of dummies available compared to a limited set controlled by the DDA.

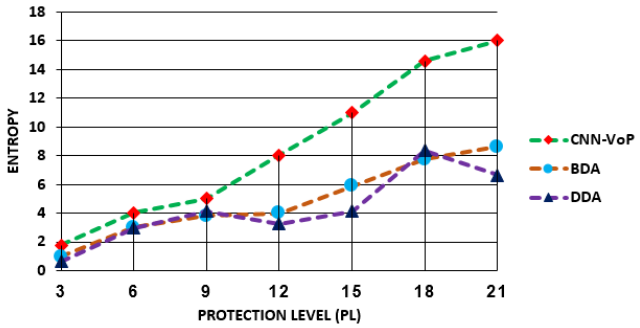

Fig. 26. Value of Entropy Metric vs. Increasing PL without any Threat.

C. Entropy-Based Evaluation under Threats

After applying the MMA threat, the value of entropy experiences a decreasing trend compared to the normal situation; the attacker has no information about the geographic map from which LBS queries are sent. This is shown in Fig. 27.

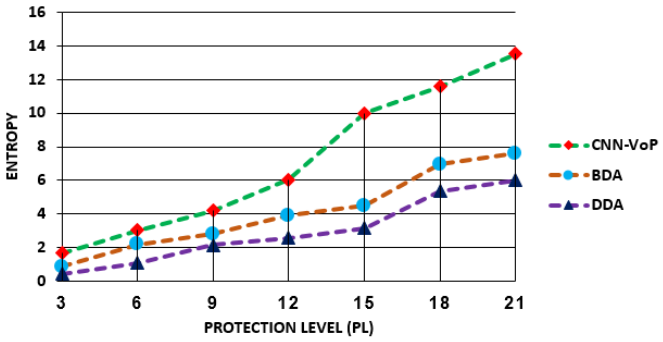

Fig. 27. Value of Entropy Metric vs. Increasing PL under MMA Threat.

Discussion and justifications: As shown in Fig. 27, despite the negative impact of the MMA threat, the CNN-VoP approach maintained its peak position. This is because of the factors taken into account in the procedure for selecting the dummies, where (1) the $CELL_p^q$ of each dummy is the same as the real location (which contributes to destroying the benefits that may be gained at the attacker side by analyzing the dummies if they are located in well-known areas) and (2) the dummies are spread over a wide space that cannot be collected in one area for malicious filtering by the attacker. The BDA scheme overcomes the DDA scheme since DDA is vulnerable to selecting the dummies based on area similarities. This gap can be exploited by the attacker to filter some dummies,

weakening the defense that is created by the DDA. In contrast, the BDA scheme ignores the similarities since it selects dummies randomly, avoiding the drawback of the DDA scheme. Table III shows the numerical results of the entropy metric.

The results summarized in Table III show that the BDA scheme experiences slight weaknesses against MMA attacks compared to a significant weakness in the DDA scheme.

For robustness against the SLA attack, Fig. 28 shows the documented results, where a severe negative impact is clearly seen in the DDA approach compared to the normal situation.

Discussion and justifications: As shown in Fig. 28, the common fact that "the value of entropy increases as the PL increases" is still working in all schemes involved in comparison. However, the entropy values are less when compared to the values under MMA threat, which reflects that the SLA is more dangerous than the MMA threat. Despite this change, the CNN-VoP scheme is still ranked at the top, followed by the BDA scheme. At the last position, the DDA scheme is coming. This scenario can be justified by the positive contribution of using CNN to scan and elect strong dummies for membership in the final set of dummies used for privacy protection. The BDA scheme ignores the factor related to elect dummies that achieve the condition of distance (i.e., long distances between the elected dummies and the real location). The DDA scheme employs none of the factors, and therefore, it performs the worst as a protection method. Table IV shows the entropy values after applying the SLA threat.

To address the difference between the MMA and SLA threats, Fig. 29 shows a visual representation of the entropy values documented in Table III and Table IV.

Table V shows the transformation of the visual representation of Fig. 29 into numeric values.

TABLE III. RESULTS OF ENTROPY IN THE THREE SCHEMES UNDER MMA THREAT

Approach \ PL		3	6	9	12	15	18	21
CNN-VoP	Entropy Values	1.658	3.046	4.208	6.046	9.987	11.587	13.508
BDA		0.854	2.196	2.809	3.906	4.501	6.946	7.609
DDA		0.427	1.078	2.145	2.578	3.150	5.347	6.005

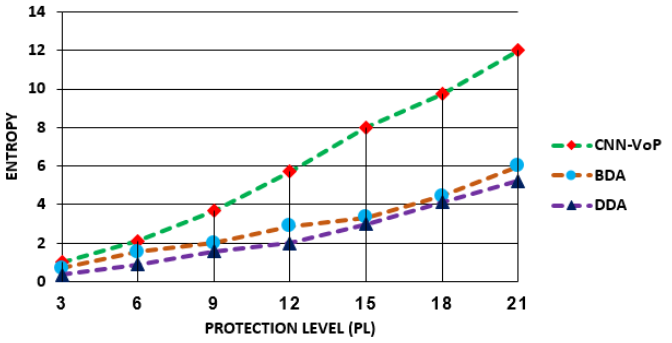

Fig. 28. Value of Entropy Metric vs. Increasing PL under SLA Threat.

TABLE IV. RESULTS OF ENTROPY IN THE THREE SCHEMES UNDER SLA THREAT

Approach \ PL		3	6	9	12	15	18	21
CNN-VoP	Entropy Values	1.007	2.113	3.666	5.711	7.999	9.720	11.994
BDA		0.700	1.539	1.999	2.878	3.332	4.448	5.996
DDA		0.364	0.886	1.589	1.988	2.997	4.123	5.231

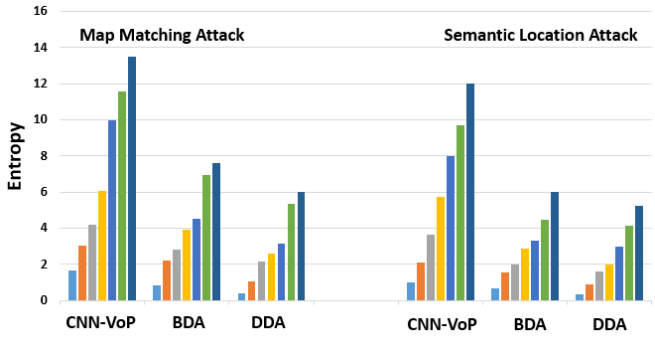

Fig. 29. Visual Representation of the Entropy values under MMA and SLA Threats.

TABLE V. DIFFERENCE IN ENTROPY VALUES AMONG THE THREE SCHEMES AFTER APPLYING MMA AND SLA THREATS

Approach \ PL		3	6	9	12	15	18	21
CNN-VoP	Difference of Entropy Values	0.651	0.933	0.542	0.335	1.988	1.867	1.514
BDA		0.154	0.657	0.81	1.028	1.169	2.498	1.613
DDA		0.063	0.192	0.556	0.579	0.153	1.224	0.774

Table V shows that the SLA threat has a more negative impact on the privacy of LBS users than the MMA threat. This is because the attacker employs time usage and knowledge about the geographic map to attack privacy (or filter some dummies). Thus, it is recommended to pay more attention to the semantic location threat in the location privacy research arena.

D. Accuracy-Based Evaluation

For the accuracy of electing suitable (or strong dummy locations), Fig. 30 illustrates the output of the three schemes.

Discussion and justifications: As shown in Fig. 30, the CNN-VoP scheme performs the best, followed by the BDA and DDA schemes. The root reason for this is related to using SVM as a classifier in the structure of the CNN used to scan and discover the dummy locations. Due to the series of convolutional and pooling layers used in the CNN, effective features of the region that includes the real location of the LBS user are generated. Based on the extracted features, strong dummies that satisfy the two conditions are elected. This means that some strong base criteria are used in the CNN-VoP scheme compared to poor ones in the other two schemes. This helps to add another strong justification about the strength of the CNN-VoP scheme in deep learning. This, in turn, provides proof of why entropy values are higher in both cases (i.e., without a threat and under the threat of attack) discussed above.

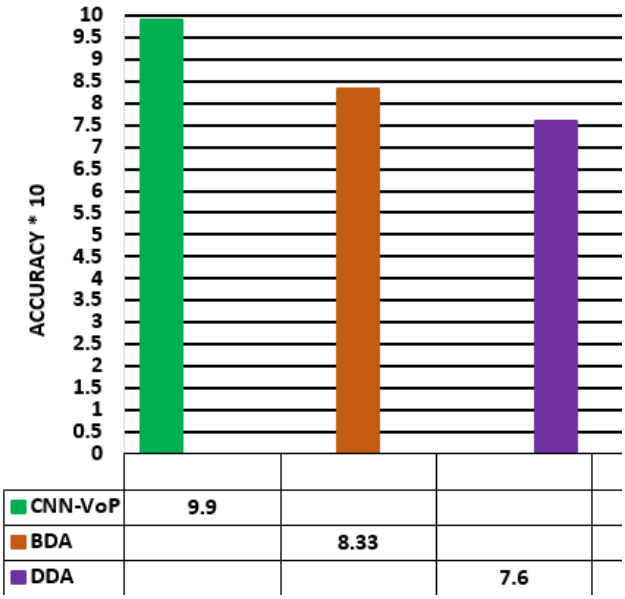

Fig. 30. Accuracy of Three Schemes

E. Performance-Based Evaluation

According to the increased number of sent Q_{LBS}, the total execution time is calculated for the three schemes, as shown in Fig. 31.

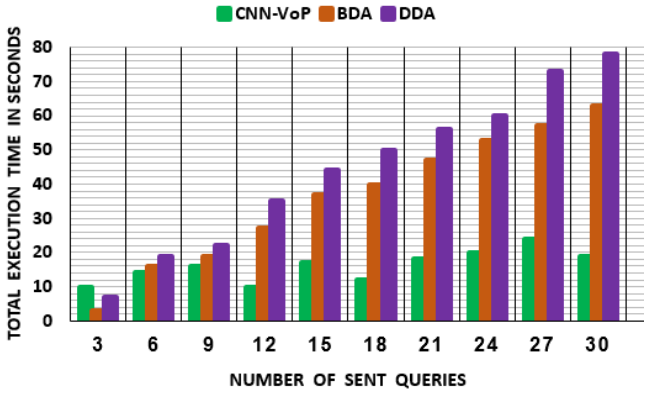

Fig. 31. Performance of Three Schemes in Terms of $TexeT$.

Discussion and justifications: As shown in Fig. 31, the CNN-VoP scheme performs the best compared to the BDA and DDA schemes, except at the beginning, where it is the worst. The reason that the CNN-VoP scheme performs the worst at the beginning is related to the training stage, where it still has no knowledge and requires much more time. During this period, the BDA scheme selects the dummies randomly, which is a fast method and leads to the shortest time. The DDA scheme consumes some time to construct the tree of dummies and thus comes after the BDA. After that, when increasing the number of sent queries, $TexeT$ increases in all schemes. However, the CNN-VoP scheme achieves the best performance since, after the training completes, it can select dummies directly based on a knowledge database. The BDA scheme returns to the second-order, while the DDA performs the worst (the longest time). This is related to a common problem when using a decision tree, which is an overfitting problem. This means that the process of constructing the decision tree requires manipulating all branches without ignoring any branches. This consumes substantial time and leads to poor performance by the DDA scheme. In terms of average, $TexeT = 17\ sec$ for the CNN-VoP scheme, $TexeT = 47\ sec$ for the BDA scheme, and $TexeT = 67\ sec$ for the DDA scheme.

VI. Conclusion and Future Work

Recently, the world witnessed a widespread COVID-19 pandemic which changed the way people performed daily tasks. In this context, and to avoid infection, people tended to use location-based services (LBSs), which have received great attention from companies and research groups. Relying on an LBS opens the door for attackers to attack the privacy of LBS users since performing tasks requires sending the user's real location. The problem is accentuated concerning advanced methods that attackers can use, such as Map Matching Attacks (MMAs) and Semantic Location Attacks (SLAs). The privacy of LBS users will be under great threat if the LBS provider acts as an attacker and can apply MMA and SLA attacks. In responding to this challenge, this work presents a location privacy protection system. The system consists of three main components. The first component is the intelligent finder. The role of the intelligent finder is to find (or select) strong dummy locations for privacy protection against the malicious party (the LBS provider), such that the attacker will be confused about determining the real location of the LBS user among the dummies. The intelligent finder uses a deep learning technique, which is the Convolutional Neural Network (CNN). The CNN is employed to create a classifier that classifies locations found in the region where the LBS user is located into the categories of weak and strong dummies. After creating the strong dummy category, a Vector of Protection (VoP) approach is performed. Strong dummies satisfy two main constraints: (1) the query probability of each selected dummy is the same as the real location, and (2) they are spread away from each other and the real location. The previous two constraints ensure high resistance against advanced MMA and SLA threats. The second component is the query builder, which is responsible for (1) constructing the protected query based on the selected strong dummies and (2) hiding the identity of the LBS user. The third component is the sender, which is responsible for sending the protected query to the LBS provider. The proposed location privacy protection system is evaluated according to entropy (the privacy protection metric), accuracy (the deep learning metric), and total execution time (the performance metric). Compared to well-known systems, which are the DDA and the BDA, the proposed system shows better results, where entropy = 15.9, accuracy = 9.9, and total execution time = 17 sec.

Limitation: Privacy protection for an LBS considers location privacy and query privacy; the sent query can be analyzed depending on the query sampling attack. In attacking query privacy, the attacker relies on the PoI as well as its link with the locations. In this work, query privacy was not taken into consideration.

Future work: In future work, we will enhance the proposed system to ensure comprehensive privacy protection in LBS

applications (i.e., ensuring both location privacy and query privacy). In addition, we will test the system using different databases for training the intelligent finder and use another advanced intelligent method, such as advanced clustering.

REFERENCES

[1] Aceto, Giuseppe, Valerio Persico, and Antonio Pescapé. "Industry 4.0 and health: Internet of things, big data, and cloud computing for healthcare 4.0." Journal of Industrial Information Integration 18 (2020): 100129.

[2] Kassab, Wafa'A., and Khalid A. Darabkh. "A–Z survey of Internet of Things: Architectures, protocols, applications, recent advances, future directions and recommendations." Journal of Network and Computer Applications 163 (2020): 102663.

[3] Jiang, Hongbo, et al. "Location Privacy-preserving Mechanisms in Location-based Services: A Comprehensive Survey." ACM Computing Surveys (CSUR) 54.1 (2021): 1-36.

[4] Garg, Niharika. "Technology in Healthcare: Vision of Smart Hospitals." Handbook of Research on Engineering, Business, and Healthcare Applications of Data Science and Analytics. IGI Global, 2021. 346-362.

[5] Ma, Yujun, et al. "Big health application system based on health internet of things and big data." IEEE Access 5 (2016): 7885-7897.

[6] Girginkaya Akdağ, Suzan, and Ahu Ergen. "Role of location-based mobile apps in city marketing: Beşiktaş as a student-friendly district." Journal of Location Based Services 14.2 (2020): 49-70.

[7] Uphaus, PerOle, et al. "Location-based services–the market: success factors and emerging trends from an exploratory approach." Journal of Location Based Services (2021): 1-26.

[8] Wawrowski, Bartosz, and Iwona Otola. "Social Media Marketing in Creative Industries: How to Use Social Media Marketing to Promote Computer Games?." Information 11.5 (2020): 242.

[9] Zou, Shihong, et al. "CrowdHB: A Decentralized Location Privacy-Preserving Crowdsensing System Based on a Hybrid Blockchain Network." IEEE Internet of Things Journal (2021).

[10] Jiang, Hongbo, et al. "Location Privacy-preserving Mechanisms in Location-based Services: A Comprehensive Survey." ACM Computing Surveys (CSUR) 54.1 (2021): 1-36.

[11] Almusaylim, Zahrah A., and N. Z. Jhanjhi. "Comprehensive review: Privacy protection of user in location-aware services of mobile cloud computing." Wireless Personal Communications 111.1 (2020): 541-564.

[12] Lv, Wenzhe, et al. "Towards Large-Scale and Privacy-Preserving Contact Tracing in COVID-19 pandemic: A Blockchain Perspective." IEEE Transactions on Network Science and Engineering (2020).

[13] Dardari, Davide, Pau Closas, and Petar M. Djurić. "Indoor tracking: Theory, methods, and technologies." IEEE Transactions on Vehicular Technology 64.4 (2015): 1263-1278.

[14] Zhang, Lan, et al. "Montage: Combine frames with movement continuity for realtime multi-user tracking." IEEE Transactions on Mobile Computing 16.4 (2017): 1019-1031.

[15] Liu, Ting, et al. "User Personalized Location k Anonymity Privacy Protection Scheme with Controllable Service Quality." International Conference on Machine Learning for Cyber Security. Springer, Cham, 2020.

[16] Jagwani, Priti, and Saroj Kaushik. "Privacy in location based services: Protection strategies, attack models and open challenges." International conference on information science and applications. Springer, Singapore, 2017.

[17] Wernke, Marius, et al. "A classification of location privacy attacks and approaches." Personal and ubiquitous computing 18.1 (2014): 163-175.

[18] Kuang, Li, et al. "Using location semantics to realize personalized road network location privacy protection." EURASIP Journal on Wireless Communications and Networking 2020.1 (2020): 1-16.

[19] Lee, Byoungyoung, et al. "Protecting location privacy using location semantics." Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. 2011.

[20] Alrahhal, Mohamad Shady, et al. "AES-Route Server Model for Location based Services in Road Networks." International Journal of Advanced Computer Science And Applications 8.8 (2017): 361-368.

[21] Lu, Hua, Christian S. Jensen, and Man Lung Yiu. "Pad: privacy-area aware, dummy-based location privacy in mobile services." In Proceedings of the Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access, pp. 16-23. ACM, 2008.

[22] Wernke, M., Skvortsov, P., Dürr, F., & Rothermel, K. (2014). A classifcation of location privacy attacks and approaches. Pers Personal and Ubiquitous Computing, 18(01), 163–175.

[23] Jiang, Nanlan, Sai Yang, and Pingping Xu. "Enabling Location Privacy Preservation in MANETs Based on Distance, Angle, and Spatial Cloaking." Electronics 9.3 (2020): 458.

[24] J.-H. Song, V. W. S. Wong, and V. C. M. Leung, “Wireless location privacy protection in vehicular Ad-Hoc networks,” Mobile Networks and Applications, vol. 15, no. 1, pp. 160–171, 2010.

[25] Grissa, Mohamed, Attila Altay Yavuz, and Bechir Hamdaoui. "Location privacy in cognitive radios with multi-server private information retrieval." IEEE Transactions on Cognitive Communications and Networking 5.4 (2019): 949-962.

[26] Wu, Zongda, et al. "Constructing dummy query sequences to protect location privacy and query privacy in location-based services." World Wide Web 24.1 (2021): 25-49.

[27] SNAP website, (2018), avaliable: https://snap.stanford.edu/data/loc-brightkite.html. (accessed on 9 Oct, 2021).

[28] Alrahhal, Mohamad Shady, Maher Khemakhem, and Kamal Jambi. "A Survey on Privacy of Location-Based Services: Classification, Inference Attacks, and Challenges." Journal of Theoretical & Applied Information Technology 95.24 (2017).

[29] Bettini, Claudio. "Privacy protection in location-based services: a survey." Handbook of Mobile Data Privacy. Springer, Cham, 2018. 73-96.

[30] Gupta, Ruchika, and Udai Pratap Rao. "An exploration to location based service and its privacy preserving techniques: a survey." Wireless Personal Communications 96.2 (2017): 1973-2007.

[31] Tefera, Mulugeta K., Xiaolong Yang, and Qifu Tyler Sun. "A Survey of System Architectures, Privacy Preservation, and Main Research Challenges on Location-Based Services." KSII Transactions on Internet & Information Systems 13.6 (2019).

[32] Rajashekar, M. B., and S. Meenakshi Sundaram. "A Survey on User’s Location Detail Privacy-Preserving Models." SN Computer Science 1 (2020): 1-6.

[33] Alrahhal, Hosam, et al. "A Symbiotic Relationship Based Leader Approach for Privacy Protection in Location Based Services." ISPRS International Journal of Geo-Information 9.6 (2020): 408.

[34] Mohamad Shady Alrahhal, Maher Khemakhem and Kamal Jambi, “Agent-Based System for Efficient kNN Query Processing with Comprehensive Privacy Protection” International Journal of Advanced Computer Science and Applications(IJACSA), 9(1), 2018. http://dx.doi.org/10.14569/IJACSA.2018.090108.

[35] Mohamad Shady Alrahhal, Muhammad Usman Ashraf, Adnan Abesen and Sabah Arif, “AES-Route Server Model for Location based Services in Road Networks” International Journal of Advanced Computer Science and Applications(IJACSA), 8(8), 2017. http://dx.doi.org/10.14569/IJACSA.2017.080847.

[36] Alrahhal, Mohamad Shady, Maher Khemekhem, and Kamal Jambi. "Achieving load balancing between privacy protection level and power consumption in location based services." (2018).

[37] Alrahhal, Mohamad Shady, and Adnan Abi Sen. "Data mining, big data, and artificial intelligence: An overview, challenges, and research questions." (2018).

[38] Mona Alfifi, Mohamad Shady Alrahhal, Samir Bataineh and Mohammad Mezher, “Enhanced Artificial Intelligence System for Diagnosing and Predicting Breast Cancer using Deep Learning” International Journal of Advanced Computer Science and Applications(IJACSA), 11(7), 2020. http://dx.doi.org/10.14569/IJACSA.2020.0110763.