analytics

MDPI

Article

A Voronoi-Based Semantically Balanced Dummy Generation Framework for Location Privacy

Aditya Tadakaluru * and Xiao Qin

Department of Computer Science and Software Engineering, Auburn University, Auburn, AL 36849, USA
* Correspondence: avt0006@auburn.edu

Abstract: Location-based services (LBS) require users to provide their current location for service delivery and customization. Location privacy protection addresses concerns associated with the potential mishandling of location information submitted to the LBS provider. Location accuracy has a direct impact on the quality of service (QoS), where higher location accuracy results in better QoS. In general, the main goal of any location privacy technique is to achieve maximum QoS while providing minimum or no location information if possible, and using dummy locations is one such location privacy technique. In this paper, we introduced a temporal constraint attack whereby an adversary can exploit the temporal constraints associated with the semantic category of locations to eliminate dummy locations and identify the true location. We demonstrated how an adversary can devise a temporal constraint attack to breach the location privacy of a residential location. We addressed this major limitation of the current dummy approaches with a novel Voronoi-based semantically balanced framework (VSBDG) capable of generating dummy locations that can withstand a temporal constraint attack. Built based on real-world geospatial datasets, the VSBDG framework leverages spatial relationships and operations. Our results show a high physical dispersion cosine similarity of 0.988 between the semantic categories even with larger location set sizes. This indicates a strong and scalable semantic balance for each semantic category within the VSBDG's output location set. The VSBDG algorithm is capable of producing location sets with high average minimum dispersion distance values of 5861.894 m for residential locations and 6258.046 m for POI locations. The findings demonstrate that the locations within each semantic category are scattered farther apart, entailing optimized location privacy.

Keywords: location privacy; dummy; semantic balance; Voronoi; spatial; temporal

Citation: Tadakaluru, A.; Qin, X. A Voronoi-Based Semantically Balanced Dummy Generation Framework for Location Privacy. *Analytics* **2023**, *2*, 246–264. https://doi.org/10.3390/analytics2010013

Academic Editors: Jesus S. Aguilar-Ruiz and Ernesto Mauro Suarez Lopez

Received: 6 January 2023
Revised: 6 February 2023
Accepted: 27 February 2023
Published: 3 March 2023

1. Introduction

The widespread usage of mobile and smart Internet of Things (IoT) devices in our daily lives has led to the universal adoption of Location Based Services (LBS) as a way to customize service offerings anchored on users' geographic locations [1]. In a typical LBS scenario, users share their current location with the LBS service provider in exchange for geographically personalized services without much control over what happens to their location information after service delivery [2]. The possibility for the exploitation of location data by an LBS provider or a data breach by an adversary opens the door for location privacy concerns. Although location sharing has become a necessity for receiving high-quality service in an LBS scenario, it is critical to achieving the maximum quality of service without sacrificing the location privacy of the users. Sending dummy locations alongside the true location to the LBS server is a well-explored approach for location privacy. The LBS server would not be able to—under a perfect scenario—distinguish true from dummy locations, so the server processes all locations and sends the results back to the users to achieve location privacy. The dummy locations approach receives highly accurate LBS query results because users' true locations are delivered, thereby offering a better quality of service compared to other location privacy techniques [3]. However, current dummy

approaches overlook the distinction between various semantic categories and their intrinsic temporal constraints such as residential versus non-residential locations. Due to these limitations, the dummy locations produced are susceptible to temporal constraint attack by an adversary in scenarios where the true location of the user is residential. This erosion of location privacy protection for users of IoT devices whose legitimate locations are residential locations urges the need for a solution that can originate dummy locations capable of thwarting temporal constraint attacks by an adversary. In this paper, a novel approach to generate dummy locations that are effective against temporal constraint attack and preserve the location privacy of residential users is proposed and developed.

1.1. Background

More often than not, mobile apps and websites that offer LBS services require the user to provide their current locations. The potential for misuse of location information by LBS providers, coupled with the concerns over a possible breach of LBS servers resulting in exposing user information to an adversary, makes a strong case for location privacy protection [3]. In many cases, the quality of LBS services is directly related to the accuracy of user locations. In other words, lowering the accuracy of the true locations lowers the quality of results returned by an LBS query. The main goal of any location privacy algorithm is to maximize quality of service while protecting user locations by sharing as little as possible or not sharing the exact locations of users. The current approaches for location privacy can be classified into four main categories—cloaking, dummy location, obfuscation, and cryptographic [3]. Cloaking and obfuscation do not send the true location of the user to the LBS server, thus resulting in a lower quality of services—specifically, LBS services that require the exact location of the user [4]. The cryptography-based approaches are computationally intensive, which renders them impractical [2]. The dummy location approaches involve sending a user's genuine location along with the dummy locations, making these preferable for achieving high-quality LBS. The two main categories of LBS services are (1) snapshot and (2) continuous services [3]. In a snapshot LBS, the user's location is submitted only once to the LBS server for query results, whereas, in a continuous LBS, user locations are continuously reported to an LBS server to receive up-to-date query results. An example of a snapshot LBS would be searching for the nearest point of interest (POI) such as restaurants or hotels. An example of continuous LBS services would be using an application for driving directions where the user's current location is continuously sent to the LBS server to track the user's movement and provide up-to-date driving directions [3]. This paper's research mainly focuses on generating dummy locations for location privacy in the context of a snapshot LBS.

1.2. Temporal Constraint Attack

In a temporal constraint attack, an adversary uses location information from historical dummy-based LBS requests delivered by a specific user and exploits the differences in temporal constraints between semantic categories, such as residential versus POI, in a historical timeline to separate the dummy locations from a true user location. By eliminating the dummies using a temporal constraint attack, an adversary is likely to identify the real location of the user breaching the location privacy of the user. The primary purpose of a residential location is housing, and it is where people live [5]. This fundamental assumption that people live in their houses and not in their workplace or a restaurant forms the basis for our argument that a residential location has different temporal constraints from point of interest (POI)-based locations. When dummy locations are generated without consideration of the distinction between residential and non-residential locations, an adversary can exploit the semantic difference to eliminate dummy locations based on temporal constraints and identify the true location of a user.

For a given location, an adversary can use approaches such as reverse geocoding [6] and other background information to find the specific address and the semantic category of the location, such as residential or supermarket. Using this technique, an adversary can retrieve semantic categories for all the locations in a dummy-based historical LBS request. The adversary can then compare and evaluate the semantic categories of all these locations in a historical timeline. By carrying out this evaluation within the context of temporal constraints associated with semantic categories such as residential versus non-residential purposes, an adversary can eliminate the dummy locations and increase the probability of true location identification. Depending on the amount of historical location requests possessed and the extent of semantic diversity implemented in generating dummy locations, an adversary tends to be able to successfully eliminate all dummy locations and to identify users' legitimate locations.

In the following example, we demonstrate how an adversary can exploit historical dummy-based LBS requests by viewing them in a temporal context of a single 24 h period beginning at midnight (12 a.m.) up to 11.59 p.m. on the same day. Table 1 displays the semantic information associated with locations in sample LBS requests containing true and dummy locations sent to the LBS server by a single user at various times over one day. Each row corresponds to a request made at a certain time of the day and shows semantic categories for five locations $\{l_1, \ldots, l_{k=5}\}$ where k is the total number of locations dispatched to the LBS server in each request, with $k-1$ dummy locations and one true location ($Residential_{True}$).

Table 1. Showing the semantic information associated with locations in sample LBS requests.

	Location Semantic Type				
Time	$\mathbf{l_1}$	$\mathbf{l_2}$	$\mathbf{l_3}$	$\mathbf{l_4}$	$\mathbf{l_5}$
2:20 a.m.	Restaurant	$\text{Residential}_{\text{True}}$	Supermarket	Shopping mall	Gas station
4:01 a.m.	Restaurant	$\text{Residential}_{\text{True}}$	Supermarket	Shopping mall	Gas station
6:10 a.m.	Restaurant	$\text{Residential}_{\text{True}}$	Supermarket	Shopping mall	Gas station
9:15 a.m.	Restaurant	$\text{Residential}_{\text{True}}$	Supermarket	Shopping mall	Gas station
11:50 a.m.	Restaurant	$\text{Residential}_{\text{True}}$	Supermarket	Shopping mall	Gas station
3:00 p.m.	Restaurant	$\text{Residential}_{\text{True}}$	Supermarket	Shopping mall	Gas station
8.00 p.m.	Restaurant	$\text{Residential}_{\text{True}}$	Supermarket	Shopping mall	Gas station
11:20 p.m.	Restaurant	$\text{Residential}_{\text{True}}$	Supermarket	Shopping mall	Gas station

In this concrete example, we assume that an adversary possesses background information that all the LBS requests listed in the table belong to the same user and request type, and occurred on the same day within 24 h between midnight (12 a.m.) and 11:59 p.m. in a specific time zone. We also assume that the dummy locations submitted for a given true location do not change over time because the same dummy algorithm is used to generate the dummy locations for every new request. Moreover, it is assumed that a POI location is a place of business and is not used for residential purposes. The main goal of an adversary in a temporal constraint attack is to eliminate the $k-1$ locations and identify the k^{th} location that is also a true location. In the above example, it is easy to deduce that a location with a semantic type, shopping mall, is a dummy location since it is unusual to be at a shopping mall at 2:20 a.m. and also to be at a shopping mall throughout the entire day. Using the same logic, we demonstrate that the two semantic types—restaurant and supermarket—can be further eliminated. Although being at a gas station at 2.20 a.m. is possible, it can be pruned as a dummy as it is unusual to be at the same gas station throughout the day. By successfully eradicating the four locations, the fifth location of the residential semantic type that remains is identified as the legitimate location. The adversary can also erase all

four business locations since it is not usual for a user to be at a place of business for an entire day, and can identify the residential location as the true location.

The example shows that, despite the maximum semantic diversity between true and dummy locations, the adversary can still exploit the semantic information from the historical requests data and identify the true location by wiping out the $k-1$ dummy locations. The current dummy approaches fail to integrate the intrinsic semantic particularity, which, in this case, is the notion of home associated with a residential location [5], and instead treat the residential location as a general semantic type without any special attention. This could result in potential dummy locations that are susceptible to temporal constraint attack as shown in the example above.

1.3. Problem Statement

The semantic type of the location can have a significant impact on the effectiveness of the dummy locations in protecting a true location from identification [7]. The conventional approach is to use semantic location diversity to identify dummy locations that are semantically different from each other and, thus, make it harder to distinguish between true and dummy locations. This approach does not work in scenarios such as when the true location is residential since the residential locations are inherently different from the non-residential POI locations in terms of the purpose of use and hours of operation. A dummy approach must handle these intrinsic differences between various semantic types to successfully protect the location privacy of the user. Otherwise, this could result in generating dummy locations that are susceptible to temporal constraint attacks. There is no known solution to this problem since the existing dummy approaches either do not acknowledge the semantic differences associated with handling a true location such as residence versus POI, or do not factor in the semantic type of the location for dummy generation altogether. In this paper, a novel dummy generation framework is devised to produce semantically balanced dummy locations that can withstand a temporal constraint attack by adversaries. Although the focus of this study is on addressing the temporal constraint attack in the case of a user whose true location is a residence, the proposed framework is capable of furnishing comprehensive location privacy for true locations of both residential and non-residential POI semantic types. To the best of our knowledge, this paper is the first among its peers to present temporal constraint attacks along with a candidate solution.

1.4. Major Contributions

The major contributions of this paper are as follows:

1. We introduce a new type of location privacy attack called 'temporal constraint attack' where an adversary can exploit the location semantics from a temporal dimension for eliminating dummies and identifying the true location. In doing so, we provide evidence on how a true location of residential semantic type can be compromised in a temporal constraint attack.
2. A novel Voronoi-based semantically balanced dummy generation (VSBDG) approach is proposed to generate dummy locations that are capable of withstanding a temporal constraint attack by an adversary. In general, the VSBDG algorithm can achieve location privacy protection regardless of the semantic type of the true location; whether it is residential or non-residential. This is due to the semantically balanced nature of the location set generated by the VSBDG.
3. One of the major drawbacks of existing dummy location studies is that they do not consider the spatial context of the location, which is not possible unless the technique is built upon real-world geospatial datasets. At best, the current approaches are tested on simple real-world location datasets that contain a collection of point locations. The VSBDG algorithm is built and tested on real-world geospatial datasets such as land parcels and point of interest (POI) locations. The VSBDG algorithm leverages spatial relationships and operations to identify spatially similar dummy locations for a given true location.

4. The Voronoi polygons are applied to model and delineate POI influence. We establish an approach that uses a cosine similarity search for finding geographical areas within the city with similar POI influence, and perform a parcel-based similarity search to identify the residential dummy location within each similar Voronoi polygon. This allowed us to identify spatially similar residential and POI dummy locations and build semantically balanced location sets that are resistant not only to temporal constraint attacks but also to location homogeneity attacks, location distribution attacks, and map-matching attacks.

The rest of the paper is organized as follows. Section 2 describes the related work. Section 3 presents the proposed methodology. In Section 4, we provide a detailed explanation of the VSBDG algorithm. Section 5 articulates the experimental implementation of the VSBDG on one sample location followed by a detailed experiments for verifying effectiveness of the VSBDG algorithm. Section 6 evaluates the results from Section 5.3. Finally, we state the conclusions and future directions in Section 7.

2. Related Work

Dummy generation techniques for location privacy in location-based services or LBS are well-studied. In general, dummy locations' techniques operate under the premise that sending dummy locations along with a true location will help conceal the true location from identification by LBS servers or adversaries [8]. Hence, for dummy locations to successfully preserve the location privacy of the true location, it is of paramount importance for the dummies to be indistinguishable from the legitimate one. Without the high degree of similarity between true locations and dummies, an adversary can exploit the dissimilarities for eliminating dummies and identifying the true location. The dummy identification process plays a vital role in ensuring that both dummy and true locations are indistinguishable from one another, thereby maximizing location privacy.

A handful of early studies spearheaded the design of dummy generation techniques. For example, Kido et al. [9] and Lu et al. [10] generated dummy locations without consideration of their similarity to true locations. Niu et al. [11] proposed V-circle and V-grid algorithms where final dummy locations are chosen based on their similarity in query probability to a genuine location. Niu et al. [12] proposed a subsequent solution that introduced an enhanced-DLS algorithm that not only generates dummy locations anchored on their similarity in location query probability but also maximizes the physical dispersion of the dummy locations to ensure good location privacy protection. Nisha et al. [13] devised a proxy-based approach in which dummies are identified from within a privacy area calculated using a proxy of the true location. The proxy instead of a true location along with the dummies is sent to the LBS server for processing. Legitimate locations are later extracted from the results received for the proxy location from the LBS server. Despite the additional privacy achieved by not disclosing true locations, the client-side extraction of true results may not be possible in real-world scenarios due to the resource constraints on the clients' devices.

Chen and Shen [7] developed MaxMinDistDS and Simp-MaxMinDistDS that determine the dummy locations using maximum semantic diversity and physical dispersion. Zhang et al. [14] presented a similar approach applying maximum semantic diversity and physical dispersion as a benchmark for selecting dummy locations. Anamala and Subramanian [15] introduced an approach to determining dummy locations using maximum semantic diversity and historical location query probabilities as criteria. Shi et al. [16] designed a dummy generation solution using semantic similarity between locations as a principle. In this approach, a semantic location is defined as a vector of historical query probabilities in a 24 h time period, during which the similarity between the semantic locations is calculated using the cosine similarity. Zhang and Li [17] advocated for a dummy generation model that combines semantic diversity, location query probabilities, and physical dispersion for producing effective dummy locations. Semantic diversity helps maximize location privacy in scenarios where genuine locations are situated in a

semantically homogeneous area. However, this strategy fails to address the possibility of a temporal constraint attack in scenarios such as a residential true location. In this case, an adversary can exploit the semantic diversity in a temporal dimension [18] to eliminate dummies and to identify the residential true location.

Alyousef et al. [19] implemented a novel approach to creating dummy locations using deep learning. In this solution, the dummies are generated through a convolutional neural network, based on their similarity to the true location in terms of location query probability and the resulting maximum physical dispersion. The key limitations of using location query probabilities in a real-world implementation are the difficulty associated with obtaining query probability data from a trusted source for all locations within an area of interest and the computational complexity associated with storing and processing these data. Jagarlapudi et al. [20] proposed a method of using a drone to assist with dummy generation. One of the major drawbacks of this approach lies in its reliance on a drone because it is impractical to have a drone linked to a user's device at all times.

The main goal of applying dummy locations for location privacy is to mask real locations from identification by an adversary, and this goal is achieved by producing dummy locations that are indistinguishable from the true locations [8]. The current dummy approaches employ a variety of factors such as semantic diversity, location query probabilities, physical dispersion [8], and spatial context [21] for evaluating locations. Except for semantic diversity, these approaches do not consider the primary purpose served by a particular location such as residential housing, commercial use, or office buildings. Dummy approaches built on semantic diversity incorporate the location's primary purpose as a way to categorize locations to assist with the dummy generation process [7]. This approach, however, fails in integrating the unique temporal aspects of a semantic category (e.g., residential locations), which can be exploited by adversaries in temporal constraint attacks.

Another major drawback of all the aforementioned approaches is that the spatial context is not factored in when qualifying dummy locations. Dummy locations generated without accounting for their similarity to real locations in a spatial context can be an easy target for location homogeneity and map-matching attacks [22] by an adversary. Amid a location homogeneity attack, both true and dummy locations are located in the same parcel area such as a university campus or hospital area, allowing for an easy inference about the general whereabouts of the user by an adversary. In a map-matching attack, an adversary eliminates dummies that are situated in natural areas such as lakes, mountains, and green areas by overlaying the dummy and real locations on a map. A map-matching attack is also a possibility in cases where the dummy locations are located in geographically dissimilar areas compared to the true location. Here is an example: A legitimate location is a residential house and all the dummies are located on roads. Hence, the spatial context plays a vital role in determining the dummy locations that provide maximum location privacy. Tadakaluru [21] proposed a parcel-based similarity scheme to create dummy locations that are similar in a spatial context to a true location. The study used real-world parcel data to assess and evaluate the spatial context of locations to forge dummy locations from the parcels that are spatially similar to the parcel of the input location [21].

3. Proposed Methodology

We propose a point of interest (POI)-based approach to producing semantically balanced dummy locations. A physical location associated with a POI is employed for delineating the influence of the POI within a geographical space through Voronoi polygons. Within each Voronoi POI influence area, the parcel-based dummy generation framework devised by Tadakaluru [21] is then deployed to create dummy locations that are spatially similar to legitimate ones. These topics are further articulated in detail in Sections 3.1–3.4.

3.1. Relationship between Geographic Location, Address, and Land Parcel

A geographic location is an exact physical place on the Earth's surface usually represented by a unique latitude and longitude pair. In a majority of the state-of-the-art location privacy studies, the term location is generally used to refer to a geographic location. An adversary can leverage the street address associated with the geographic location to obtain sensitive information about a user [23] and, hence, this trick plays an important role in building privacy-preserving mechanisms. An address is generally associated with a parcel of land that has designated property ownership boundaries. When someone refers to a specific geographic location (i.e., longitude, latitude) in an urban setting, it is normally linked to an address that is representative of a land parcel. In other words, any physical location located within a land parcel is associated with a unique address assigned to the land parcel [24]. Despite the importance of address and land parcels, most dummy-location generation approaches in location privacy assume that a single physical location is enough to represent an address and the underlying land parcel. As a result, these solutions use the physical location as a sole spatial component within their algorithms. In this study, we make use of land parcels rather than single locations to select areas that are similar in a spatial context to the area containing real locations. Applying the land parcels, we can delineate the geographical boundary of a POI location to support the distinction between residential and non-residential locations in the VSBDG algorithm. Without this delineation offered by the land parcel, it is impossible to guarantee that a chosen location is of a certain semantic type—an important requirement for building a semantically balanced dummy location set.

3.2. Modeling POI Influence Using Voronoi Polygons

A Voronoi diagram anchored on POI locations is employed to divide a geographical area into polygons, with each enclosing a single POI location such that any location within a given polygon is closer to the related POI than to any other POI locations [25]. The locations lying on the edge of a Voronoi polygon are equidistant to POIs associated with the two Voronoi polygons sharing an edge. The main purpose for adopting Voronoi polygons in the proposed approach is to avoid dynamic runtime analysis of POIs in such a way that an efficient dummy generation process becomes viable. This idea is made possible by the intrinsic property of a Voronoi polygon, which contains only a single POI. This novel design facilitates the selection of deterministic proportions of POIs and residential locations within a geographic area influenced by the given POI, a Voronoi polygon. The Voronoi polygons are generated as part of data pre-processing and reused for each new dummy generation request. The predictability of POIs within a group of Voronoi polygons eliminated a pressing need for proximity queries looking for the POIs within a geographical area influenced by a POI during runtime. Given the Voronoi polygons for a POI dataset, a semantically balanced location set L containing both true and dummy locations can be formally expressed as:

$L = \{l_1, l_2, \ldots, l_m, l_{m+1}, l_{m+2}, ..., l_n\}$, where
n: total number of locations,
m: number of non-residential locations (POIs),
$n - m$: number of residential locations.

- m is defined as the number of Voronoi polygons used in generating a dummy location set because each Voronoi polygon contains one POI location.
- With at least two locations selected from each Voronoi polygon, there are $2m + 1$ minimum number of dummy locations and one legitimate location.

3.3. Cosine Similarity between Voronoi Polygons

In this study, we advocate for cosine similarity to find spaces that are similar to the Voronoi polygon embracing users' genuine locations. The cosine similarity between two vectors A and B is measured using the cosine angle between the vectors, which can be calculated in the Euclidean space [26] using the following formula:

$$\text{Cosine Similarity } (A,\ B) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}. \quad (1)$$

The cosine similarity gauges similarity based on the direction of vectors using the cosine angle instead of the magnitude of vectors. This measure aligns with the goal to identify Voronoi polygons that are similar to the input Voronoi polygon in POI influence in a spatial context rather than the magnitudes of feature attributes. For this reason, the attribute vectors of both target and candidate Voronoi polygons are compared and ranked using cosine similarity ordered with highly similar polygons at the top and least similar at the bottom. Dummy locations are picked from the Voronoi polygons that are most similar in cosine relationship to the input Voronoi polygon containing the true location.

3.4. Parcel-based Similarity Search

The parcel-based location privacy framework proposed by Tadakaluru [21] uses the similarity search to select dummy locations that are similar in a spatial context to real locations. In this study, the parcel-based similarity search is deployed to forge residential dummy locations within each candidate Voronoi polygon that is cosine-similar to the Voronoi polygon containing the input location. The parcel-based similarity search is driven by the Euclidean distance between the attribute values of target and candidate parcels. For each target and candidate parcel, the total sum of squared differences (SSD) between standardized attribute values is calculated. The candidate parcels are then ranked based on their SSD values with the target parcel, where the candidate parcel with the lowest SSD value ranked higher is considered to be the most similar one to the input parcel. The SSD between two parcels P and Q with n attributes can be calculated as follows.

$$\text{SSD } (P,\ Q) = \sum_{i=1}^{n} (P_i - Q_i)^2$$

4. Voronoi-Based Semantically Balanced Dummy Generation (VSBDG)

The main objective of the proposed approach is to generate semantically balanced dummy locations that can effectively withstand a temporal constraint attack, thereby maximizing location privacy for users. The semantic balance between residential and non-residential dummy locations is accomplished by dividing a geographical area, such as a city or county bounding a true location, into separate regions based on POI influence. The Voronoi polygons are generated using the POI dataset, where each Voronoi polygon is associated with a single POI location (see also Section 3.2). The key rationale behind the deployment of Voronoi is to facilitate a guaranteed and predictable selection of one POI within each Voronoi polygon, control the ratio of residential versus non-residential (POI) dummy locations, and preserve the semantic balance of a given location set.

Algorithm 1 originates a semantically balanced location set for a given legitimate location using land parcels, POI-based Voronoi polygons, and POIs' dataset(s) for a geographical region like a city or a county. The first steps in rows 1 and 2 identify the relevant land parcel p_{true} and Voronoi polygon v_{true} outlining the true location l_t using spatial join. Then, the next major step involves using a cosine similarity search to identify the top m Voronoi polygons that are similar to v_{true}. A next similarity search based on the Euclidean distance between attributes is performed to pinpoint the land parcel ($parcels\_similar_i$) that is most similar to p_{true} in each of the m Voronoi polygons. Finally, the associated POI and the centroid of $parcels\_similar_i$ for each of m Voronoi polygons are appended to the dummy location set, returning $2m + 1$ dummy locations in total.

Algorithm 1: VSBDG—To identify semantically balanced dummy locations for a given residential true location

Input: True location l_t (Longitude, Latitude), Location set size k
Datasets: Land Parcels P, POI-based Voronoi polygons dataset V
Output: Location set D of size k
1. Determine land parcel p_{true} outlining l_t using spatial join between l_t and P
2. Determine Voronoi polygon v_{true} outlining l_t using spatial join between l_t and V
3. $m = (k-2)/2$ // m is total number of similar Voronoi polygons to be identified
4. $V_{similar}$ = Cosine Similarity Search(target = v_{true}, candidate set = V, output length = m)
// Perform cosine similarity search to identify top m Voronoi polygons from V that are similar to v_{true}
5. For each Voronoi polygon v_i in $V_{similar}$
6. Set $candidate\_parcels\_set_i$ = parcels within Voronoi polygon v_i
7. $prclSim_i$ = Euclidean Similarity Search(target = p_{true}, candidate set = $candidate\_parcels\_set_i$, output length = 1)
// Perform Euclidean Similarity Search to identify top *1* land parcel from $candidate\_parcels\_set_i$ that is -
// most similar to p_{true}
8. Calculate $dummy_{residential}$ using parcel centroid of $prclSim_i$
9. $dummy_{poi}$ = POI location associated with v_i
10. Add $dummy_{residential}$, $dummy_{poi}$ to D
11. $dummy_{vt}$ = POI location associated with v_{true}
12. Add $dummy_{vt}$ to D
13. Add l_t to D
14. Return D

The semantic composition of the location set from Algorithm 1, containing both true and dummy locations, consists of both residential and non-residential locations regardless of the semantic classification of the real location. This intention ensures that an adversary is unable to single out the real location by exploiting the background information associated with a temporal constraint, such as general operating hours for a certain type of POI [27]. The example in Table 2 illustrates the possible residential versus POI semantic composition of a location set that contains one true location and six (k = 5) dummy locations over a period of one day, like the example shown in Table 1.

Table 2. Showing the semantic information associated with locations in sample LBS requests.

	Location Semantic Type					
Time	l_1	l_2	l_3	l_4	l_5	l_6
2:20 a.m.	Residential	Residential	$\text{Residential}_{\text{True}}$	POI	POI	POI
4:01 a.m.	Residential	Residential	$\text{Residential}_{\text{True}}$	POI	POI	POI
6:10 a.m.	Residential	Residential	$\text{Residential}_{\text{True}}$	POI	POI	POI
9:15 a.m.	Residential	Residential	$\text{Residential}_{\text{True}}$	POI	POI	POI
11:50 a.m.	Residential	Residential	$\text{Residential}_{\text{True}}$	POI	POI	POI
3:00 p.m.	Residential	Residential	$\text{Residential}_{\text{True}}$	POI	POI	POI
8.00 p.m.	Residential	Residential	$\text{Residential}_{\text{True}}$	POI	POI	POI
11:20p.m.	Residential	Residential	$\text{Residential}_{\text{True}}$	POI	POI	POI

Our algorithm provides location privacy for the true location at three different tiers as shown in Figure 1. First, by consistently generating the same number of residential and POI dummy locations each time, it decreases the chance to single out the real location using a temporal constraint attack by an adversary. Second, the POI dummies are identified based on similarity in POI influence related to the true location through the use of Voronoi polygons. Third, the residential-type dummies are identified based on their parcel similarity to the parcel outlining the true location using a parcel-based similarity search. Whether the genuine location is a residential location or a POI, it is arduous for an adversary to differentiate a particular location as a true location: for any given location set, there are at least one-half of the locations that are similar to the real location, while the other half of the locations are similar to each other. The following example, revisiting the scenario presented in Section 1.2, evaluates how the VSBDG addresses the temporal constraint attack in that scenario.

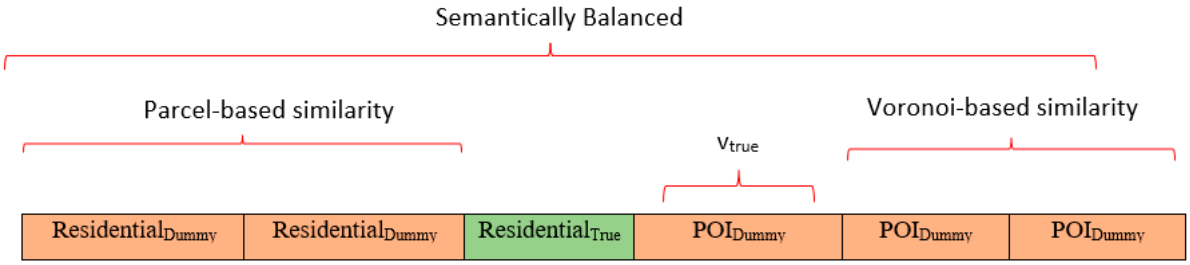

Figure 1. A three-tiered location privacy protection for a location set that contains one legitimate location and five dummy ones ($k = 6$).

Table 2 illustrates the semantic information associated with an example location set containing both true and dummy locations generated and submitted to an LBS server at various times during a single 24 h period beginning at midnight (12 a.m.) up to 11.59 p.m. on the same day. Even if an adversary employs a temporal constraint attack similar to the scenario described in Section 1.2 and identifies all the three POIs as dummies, there are still two more residential locations to identify and eliminate for the true location. The two dummy residential locations, being similar in a spatial context to the real one, are generated using the parcel-based similarity search that is proven to be effective against location homogeneity attacks and less prone to map-matching attacks [21]. By leveraging Voronoi and parcel-based similarity searches, the VSBDG algorithm generates a semantically balanced dummy location set that effectively withstands temporal constraint attack while preserving the indistinguishability of real locations.

5. Experimental Analysis and Results

5.1. Data Collection and Preprocessing

In the empirical study, we test land parcels [28] and POI datasets [29] for spatial analysis and generation of dummy locations. These datasets are gleaned for the Richmond County (Staten Island) in the state of New York, the USA. We run the geoprocessing toolkit in ArcGIS Pro [30] to extract county-level parcel and POI data from statewide New York datasets and for the rest of the analysis in this section. The parcels dataset is comprised of parcel features that are stored as polygon features, and the POI dataset contains POIs that are point features, as shown in Figure 2a. The total number of land parcels and POI datasets of the Richmond County (Staten Island) are 123,849 and 1288, respectively [31]. As stated in Section 3.2, the Voronoi polygons are generated for 1288 POI locations during preprocessing and referred to within the proposed algorithm as candidate set V.

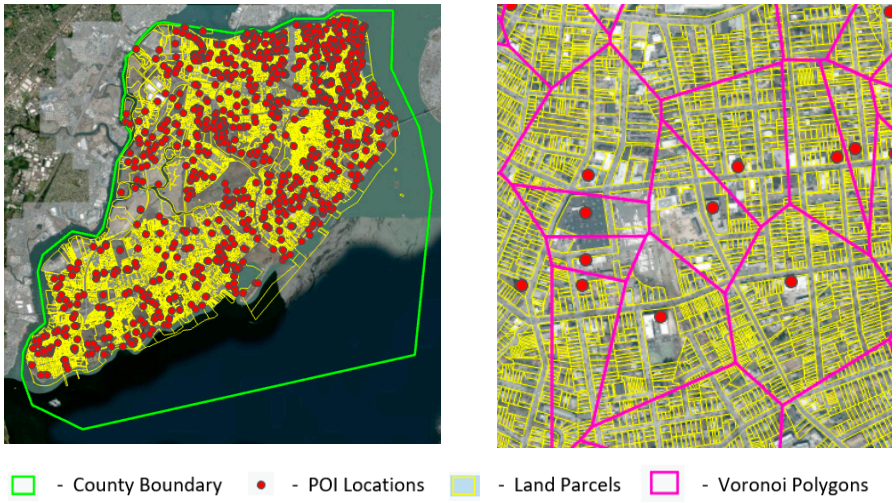

Figure 2. (**a**) Land parcels and POI locations within the Richmond County (Staten Island) overlaid on imagery basemap [32] (**b**) Voronoi polygons and their associated POIs within a section of Richmond County.

5.2. Electing Dummy Locations Using VSBDG

This section demonstrates a step-by-step implementation of the VSBDG algorithm for a sample input true location l_t and location size ($k = 2$). The first steps in the algorithm are to determine parcel p_{true} and Voronoi polygon v_{true} outlining a true location, as shown in Figure 3. The next step is to perform cosine similarity search and find the top m Voronoi polygons similar to v_{true}. The m is calculated as half of the $k - 1$ value, where k is the number of dummy locations to be created. This step is to achieve semantic balance to ensure that, for every POI dummy location, there is a residential dummy location chosen.

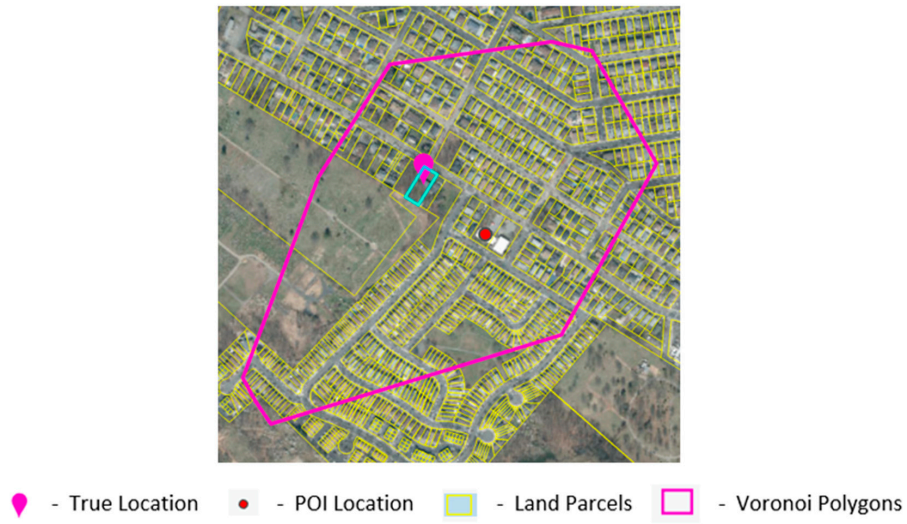

Figure 3. Input parcel p_{true} outlining the true location (highlighted in blue) and the Voronoi polygon v_{true} containing the input true location l_t as indicated in steps 1 and 2 of VSBDG (Algorithm 1).

The next step in the algorithm is to determine the top m Voronoi polygons similar to v_{true} from the candidate set V generated during the preprocessing phase. As delineated in Section 3.3, the search for similar Voronoi polygons is driven by cosine similarity and uses the attributes' area, length, and the number of land parcels within the Voronoi polygon to perform the cosine similarity search. The cosine similarity search is executed for two

dummy locations ($k = 2$) and the output for the top one similar Voronoi polygon v_1 ($m = 1$) is shown in Figure 4. The next steps involve identifying the residential parcel that is the most similar to the input parcel of the real location p_{true} within each of the m similar Voronoi polygons. Figure 4 shows a residential parcel, $prclSim_1$, that is most identical to p_{true} from all the candidate parcels within v_1 identified by the Euclidean similarity search for similar parcels. A residential dummy location, $dummy_{residential}$, is calculated using the centroid of $prclSim_1$ and the POI location linked with v_1 is chosen as the non-residential POI dummy $dummy_{poi}$. As explained in the algorithm, Steps 5 to 10 involving parcel similarity search are repeated for each of the m similar Voronoi polygons identified using cosine similarity search in the previous step. The POI location $dummy_{vt}$ associated with v_{true} is included as one of the dummies. Thus, two dummies constructed from each of m Voronoi polygon and one additional dummy, $dummy_{vt}$, results in total $2m + 1$ dummy locations for a given genuine location p_{true}. With $2m + 1$ dummy locations evaluating to three dummy locations, there are a total of four locations ($k = 4$) including the true location in the location set, as shown in Figure 5. Figure 5a depicts all the four locations in the context of Voronoi polygons, and Figure 5b shows only the locations.

Figure 4. Voronoi parcel similar to v_{true} from cosine similarity search, residential parcel $prclSim_1$ similar to p_{true} Euclidean similarity search and the two dummies ($dummy_{residential}$ and $dummy_{poi}$) identified.

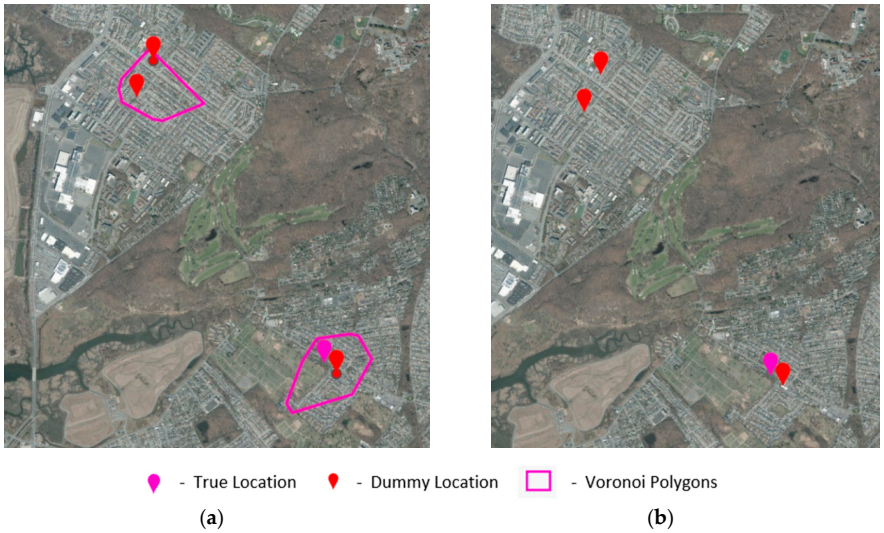

Figure 5. For a location set with four locations ($k = 4$): (**a**) shows the true location and three dummy locations with their respective Voronoi polygons and (**b**) shows the true and dummy locations only.

5.3. Results

Physical dispersion of dummy locations in a location set is employed to evaluate the effectiveness achieved by the location privacy algorithms [7,14]. The physical dispersion is measured as the minimum distance between any two locations in a location set containing both true and dummy locations [8]. A location set with a higher physical dispersion indicates that locations are scattered much farther, implying better location privacy. The core idea for building a semantically balanced location set is to ensure that, for every residential location, there exists a POI location—and vice-versa—to reduce the probability of being identified by adversaries. This goal is achieved through Voronoi polygons by choosing a set of POI and residential dummy locations from within multiple geographical areas that are spatially similar to the Voronoi area of real locations.

For a semantically balanced location set to be effective, dummy locations should not only have a higher physical dispersion within each category but should also have a physical dispersion that is similar to the other semantic category. To gauge the effectiveness of the semantically balanced location set generated by VSBDG (Algorithm 1), we bring forth an algorithm on three input true locations selected using random sampling with different location set sizes. The *VSBDG (True location l_t, Location set size k)* algorithm is implemented for three input true locations with location sizes (k) ranging from 4 to 22. Given each sample input's true location, 10 location sets containing both true and dummy locations are generated with sizes ranging from $k = 4$ to $k = 22$. For each location set, the minimum dispersion distance (MDD) is separately calculated for locations in residential and semantic categories. Table 3 tabulates the physical dispersion of residential locations for 10 location sets generated for the input legitimate location in three columns—T-RES-1, T-RES-2, and T-RES-3. Table 4 shows the physical dispersion of POI locations for 10 location sets generated for input true location in three columns—T-POI-1, T-POI-2, and T-POI-3. The minimum dispersion distance for each location set is plotted in RStudio [33] against location set size (k) separately for residential and POI locations, as shown in Figures 6 and 7.

Table 3. Showing physical dispersion of residential locations in a location set of different sizes (k) of the three input locations.

	Minimum Dispersion Distance—Residential (Meters)		
Location Set Size (k)	**True Location-1 (T-RES-1)**	**True Location-2 (T-RES-2)**	**True Location-3 (T-RES-3)**
4	9291.308514	13,685.24076	26,978.01745
6	9291.308514	10,923.37298	11,920.92343
8	9291.308514	2073.75585	11,834.15819
10	9291.308514	2073.75585	6944.580188
12	4076.586856	2073.75585	6944.580188
14	4076.586856	2073.75585	3461.144276
16	3194.775997	2073.75585	3461.144276
18	3101.138163	2073.75585	2069.387356
20	3101.138163	1618.185695	2069.387356
22	3101.138163	1618.185695	2069.387356

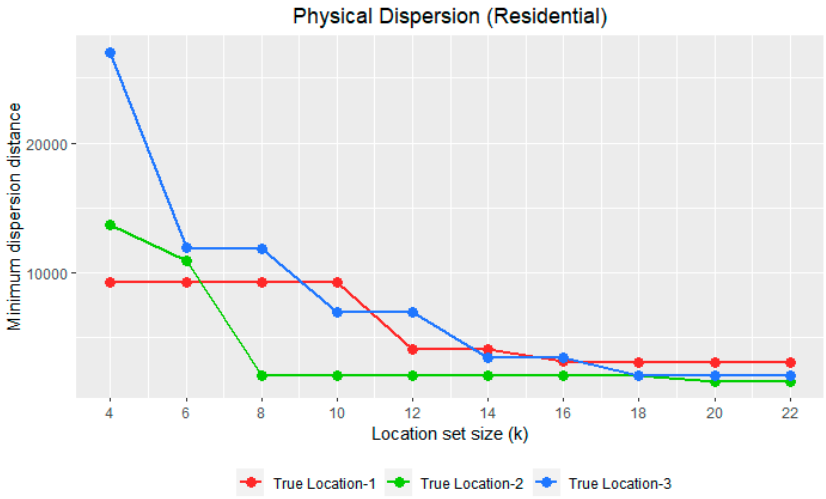

Figure 6. Plots show physical dispersion of residential locations in a location set with the size of location set (*k*) on X-axis and minimum dispersion distance (meters) on the Y-axis.

Table 4. Physical dispersion of POI locations in a location set of different sizes (*k*) for the three input locations.

	Minimum Dispersion Distance—POI (Meters)		
Location Set Size (*k*)	**True Location-1 (T-POI-1)**	**True Location-2 (T-POI-2)**	**True Location-3 (T-POI-3)**
4	10,368.38844	14,322.61609	25,800.26266
6	10,368.38844	8369.51998	12,034.79693
8	10,368.38844	3641.198031	11,580.28315
10	10,368.38844	3641.198031	7445.09195
12	3427.856197	3641.198031	7248.846177
14	3427.856197	3641.198031	3037.160013
16	3427.856197	3641.198031	3037.160013
18	2799.307792	3310.462408	2191.07635
20	2799.307792	3310.462408	2191.07635
22	2799.307792	3310.462408	2191.07635

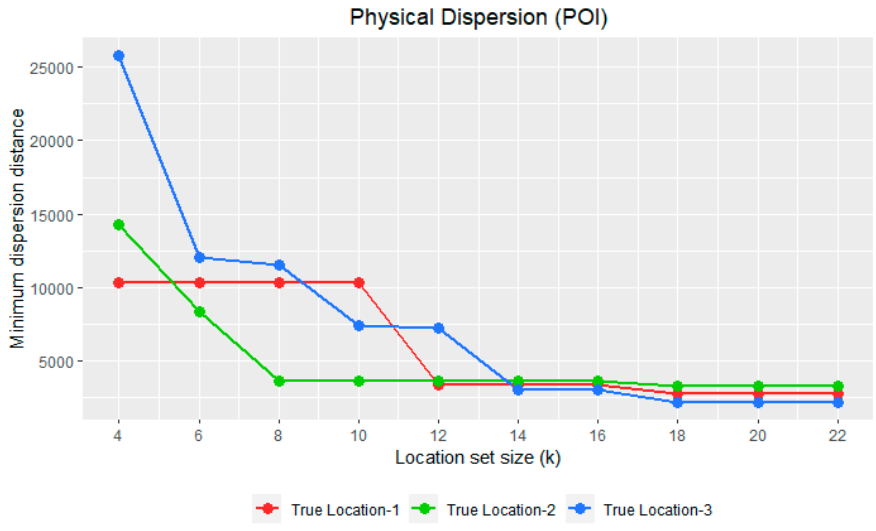

Figure 7. Plots show physical dispersion of POI locations within a location set with the size of location set (*k*) on X-axis and minimum dispersion distance (meters) on the Y-axis.

The effectiveness of a semantically balanced location set is evaluated by measuring physical dispersion similarity between residential and POI MDD values at various location set sizes, as shown in Tables 3 and 4. For this, the cosine similarity is calculated between the MDD values of residential and POI categories of each of the three input true locations separately.

6. Discussions

6.1. Evaluating VSBDG

The MDD of residential locations in the 10 location sets with sizes ranging from 4 to 22 are almost similar and follow a nearly identical trend to that of the MDD of POI locations in the same 10 location sets. This trend holds true for all the three sample locations plotted using red, green, and blue colors in Figures 6 and 7. The same is empirically proven by measuring the cosine similarity between the MDD values of residential locations versus POI locations in 10 location sets for all three input locations, as shown in Table 5. The PDCS measures of input locations 1, 2, and 3 are 0.997, 0.994, and 0.99, respectively, with an average PDCS value of 0.988, indicating a high cosine similarity between the MDD of residential locations versus the POI locations in the 10 location sets with sizes ranging from 4 to 22.

Table 5. Cosine similarity is measured between residential and POI semantic categories for the three input true locations.

Input True Location	Vectors Measured	Physical Dispersion Cosine Similarity (PDCS)
1	T-RES-1 and T-POI-1	0.9966669
2	T-RES-2 and T-POI-2	0.9641056
3	T-RES-3 and T-POI-3	0.9993439

The high average PDCS value of 0.988 between the MDD values of residential and POI locations indicates the effectiveness of the VSBDG algorithm in generating a semantically balanced location set. This high cosine similarity between the two semantic categories also demonstrates a strong semantic balance between the semantic categories that is consistent even at higher values of k. This result also unveils that VSBDG is capable of achieving a scalable semantic balance even at high values of k by equally creating efficient dummy locations, despite an increase in the size of the location set (k). This finding further confirms that our proposed algorithm is adroit at originating dummy locations that are consistent with their resistance to temporal constraint attacks despite the increase in the size of the location set k.

Without a high physical dispersion, the dummy locations are prone to location distribution attacks [18] where, in this case, an adversary can target locations from a specific semantic category. The adversary employs techniques such as clustering for eliminating dummies to identify either true locations or the neighborhood area of the true location. The latter would pose a much higher risk in a case where the real location is residential since the adversary infers a significant volume of background information by knowing the neighborhood area associated with clusters where a residential user resides [34]. The average MDD of residential locations for all the 10 location sets for each input location shown in Table 3 is 5861.894 m. The average MDD of POI locations for all the 10 location sets for each input location listed in Table 4 is 6258.046 m. These high MDD values within each semantic category demonstrate that the locations within each category are scattered farther apart, indicating optimized location privacy.

6.2. Comparison with the Existing Dummy Approaches

Recall that, to the best of our knowledge, this study is the first of its kind that introduces temporal constraint attacks, which is tackled by the VSBDG algorithm. Because of the novelty of VSBDG, it is not feasible to conduct a direct comparison of our results with relatable results produced by the other state-of-the-art dummy generation approaches. Nevertheless, we perform a comparison between VSBDG and the other state-of-the-art solutions by summarizing their handling of known location privacy vulnerabilities and features. Table 6 compares the handling of three location privacy attacks by VSBDG and the other existing state-of-the-art dummy algorithms in general. Our proposed VSBDG approach successfully handles all the three vulnerabilities listed in Table 6.

Table 6. A comparison of the proposed (VSB) and the existing dummy approaches based on how various vulnerabilities are addressed.

Vulnerability	VSBDG	COSA [21]	k-LPP [14]	VLBS [16]	DLSS [13]	V-Cir/V-grid [11]	DLIP [17]
Location homogeneity attack	✔	✔	✔	$✔_p$	X	X	X
Map-matching attack	$✔_p$	$✔_p$	X	X	X	X	X
Temporal constraint attack	✔	X	X	X	X	X	X

✔—Addresses $✔_p$—Partially addresses X—Fails to address.

The VSBDG algorithm utilizes the parcel-based similarity search from COSA [21] to seek dummy locations from parcels that are spatially similar to the parcel of an input location. Both VSBDG and COSA [21] are built on real-world geospatial datasets and leverage spatial context for dummy identification. This idea not only helps to generate dummy locations that are resistant to location homogeneity attacks but also makes them less prone to map-matching attacks [21]. The k-LPP [14], VLBS [16], and DLIP [17] address location homogeneity attacks through dummy generation based on semantic diversity. The other three approaches [11,13] neither use spatial context nor semantic diversity, making them prone to location homogeneity attacks. Further approaches [9,11,13,14,16] do not consider the spatial context in dummy generation, making them prone to map-matching attacks. To our knowledge, the VSBDG is the only framework that is capable of addressing temporal constraint attacks by using that semantically balanced location set. Table 7 provides a comparison of general features between VSBDG and the other existing dummy approaches. Overall, the VSBDG algorithm not only addresses all the three location privacy attacks (Table 6) but also offers the key benefits (Table 7) when compared to the existing dummy approaches.

Table 7. A comparison of benefits addressed by the proposed (VSB) and the existing dummy approaches.

Key Benefits	VSBDG	COSA [21]	k-LPP [14]	VLBS [16]	DLSS [13]	V-Cir/V-grid [11]	Random [9]
Physical dispersion semantic similarity for larger k values	✔	X	X	X	X	X	X
Do not use location query probability	✔	✔	✔	✔	X	X	✔
Use spatial context in dummy identification process	✔	✔	X	X	X	X	X
Do not submit proxy instead of true location to LBS server	✔	✔	✔	✔	X	✔	✔
Built on real-world geospatial dataset(s)	✔	✔	X	X	X	X	X

✔—YES; X—NO.

The potential applications of VSBDG include location privacy scenarios where a semantic category is inherently different from the other semantic categories, and one example is residential locations. In the case of residential locations, the VSBDG algorithm is leveraged to protect the real locations of a user whose locations belong to a residence. The smart devices located in a residential location are also potential candidates for location privacy protection offered by VSBDG.

7. Conclusions

Locations are unique and may differ in characteristics from locations of the other semantic types. One such case is residential locations that are unique and different in temporal constraints from locations of the POI semantic types. A dummy generation approach ought to address these intrinsic differences by constructing robust dummies that are expected to effectively conceal true locations and secure their privacy. In this paper, we identified and explored a new type of attack called a temporal constraint attack, in which an adversary exploits differences in temporal constraints between locations of different semantic types to eliminate dummy locations and single out real locations. We demonstrated how residential locations are susceptible to temporal constraint attacks when an adversary possesses historical request data on dummies submitted for a residential location. The existing techniques, including the ones that are built based on semantic diversity, are prone to temporal constraint attacks because the difference in temporal constraints of a semantic category such as residential location is not taken into account. The key takeaways from this study are summarized below.

- We proposed a novel VSBDG algorithm, which is conducive to generating dummies that can keep temporal constraint attacks at bay.
- The VSBDG algorithm is capable of handling both location homogeneity attacks and map-matching attacks for two reasons. 1. POI influence in a spatial area is modeled using Voronoi polygons and leverages cosine similarity search to find areas within a city that has similar POI influence. 2. The parcel-based similarity search [21] is adopted to construct dummy locations within each Voronoi polygon from parcels that are spatially similar to a legitimate location's parcel.
- Our findings show a high average MDD of 5861.894 m and 6258.046 m for residential and POI locations, respectively, entailing that the locations are distributed further apart indicating optimized location privacy.
- The results unfold an average PDCS of 0.988 between the MDD values of residential and POI locations in location sets with sizes ranging from 4 to 22, thereby demonstrating a strong and scalable semantic balance within an output location set of the VSBDG algorithm, suggesting good location privacy protection against a temporal constraint attack.

The temporal constraint attack model discussed in this study, accompanied by the proposed VSBDG algorithm, is specific to snapshot LBS scenarios involving a single real location. On the other hand, a continuous LBS request involves a trajectory with a series of locations [3]; hence, the temporal constraint attack in a continuous LBS scenario should be explored in a future study. Since this investigation is the first study that addresses the concerns of temporal constraint attacks, this work will pave the way for further research into the applicability of temporal constraint attacks under new scenarios and potential ground-breaking solutions addressing the temporal constraint attacks.

Author Contributions: Conceptualization, A.T. and X.Q.; Methodology, A.T. and X.Q. Formal analysis, A.T.; Investigation, A.T.; Visualization, A.T.; Writing—original draft, A.T.; Writing—review and editing, A.T. and X.Q.; Supervision, X.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sun, G.; Chang, V.; Ramachandran, M.; Sun, Z.; Li, G.; Yu, H.; Liao, D. Efficient location privacy algorithm for Internet of Things (IoT) services and applications. *J. Netw. Comput. Appl.* **2017**, *89*, 3–13. [CrossRef]
2. Liu, B.; Zhou, W.; Zhu, T.; Gao, L.; Xiang, Y. Location privacy and its applications: A systematic study. *IEEE Access* **2018**, *6*, 17606–17624. [CrossRef]
3. Jiang, H.; Li, J.; Zhao, P.; Zeng, F.; Xiao, Z.; Iyengar, A. Location Privacy-preserving Mechanisms in Location-based Services: A Comprehensive Survey. *ACM Comput. Surv.* **2021**, *54*, 1–36. [CrossRef]
4. Xu, X.; Chen, H.; Xie, L. A Location Privacy Preservation Method Based on Dummy Locations in Internet of Vehicles. *Appl. Sci.* **2021**, *11*, 4594. [CrossRef]
5. Schirmer, P.M.; van Eggermond, M.A.; Axhausen, K.W. The Role of Location in Residential Location Choice Models: A Review of Literature. *J. Transp. Land Use* **2014**, *7*, 3–21. Available online: http://www.jstor.org/stable/26202678 (accessed on 1 January 2023). [CrossRef]
6. Kounadi, O.; Lampoltshammer, T.J.; Leitner, M.; Heistracher, T. Accuracy and privacy aspects in free online reverse geocoding services. *Cartogr. Geogr. Inf. Sci.* **2013**, *40*, 140–153. [CrossRef]
7. Chen, S.H.; Shen, H. Semantic-Aware Dummy Selection for Location Privacy Preservation. In Proceedings of the 2016 IEEE Trustcom/BigDataSE/ISPA, Tianjin, China, 23–26 August 2016; pp. 752–759.
8. Zhang, S.; Li, M.; Liang, W.; Sandor, V.K.A.; Li, X. A Survey of Dummy-Based Location Privacy Protection Techniques for Location-Based Services. *Sensors* **2022**, *22*, 6141. [CrossRef] [PubMed]
9. Kido, H.; Yanagisawa, Y.; Satoh, T. Protection of Location Privacy Using Dummies for Location-Based Services. In Proceedings of the International Conference on Data Engineering Workshops, Tokyo, Japan, 3–4 April 2005. [CrossRef]
10. Lu, H.; Jensen, C.S.; Yiu, M.L. 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, Vancouver, BC, Canada, 13 June 2008.
11. Niu, B.; Zhang, Z.; Li, X.; Li, H. Privacy-Area Aware Dummy Generation Algorithms for Location-Based Services. In Proceedings of the IEEE International Conference on Communications (ICC), Sydney, Australia, 10–14 June 2014; pp. 957–962. [CrossRef]
12. Niu, B.; Li, Q.; Zhu, X.; Cao, G.; Li, H. Achieving k-Anonymity in Privacy-Aware Location-BASED services. In Proceedings of the IEEE INFOCOM 2014—IEEE Conference on Computer Communications, Toronto, ON, Canada, 27 April–2 May 2014; pp. 754–762. [CrossRef]
13. Nisha, N.; Natgunanathan, I.; Xiang, Y. An enhanced location scattering based privacy protection scheme. *IEEE Access* **2022**, *10*, 21250–21263. [CrossRef]
14. Zhang, Y.; Zhang, Q.; Li, Z.; Yan, Y.; Zhang, M. A k-anonymous Location Privacy Protection Method of Dummy Based on Geographical Semantics. *Int. J. Netw. Secur.* **2019**, *21*, 937–946.
15. Anamala, B.M.; Subramanian, S. Dispersed dummy selection approach for location-based services to preempt user-profiling. *Concurr. Comput. Pract. Exp.* **2021**, *33*, e6361. [CrossRef]
16. Shi, X.; Zhang, J.; Gong, Y. A Dummy Location Generation Algorithm Based on the Semantic Quantification of Location. In Proceedings of the IEEE International Conference Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 28–30 June 2021; pp. 172–176. [CrossRef]
17. Zhang, A.; Li, X. Research on privacy protection of dummy location interference for Location-Based Service location. *Int. J. Distrib. Sens. Netw.* **2022**, *18*, 15501329221125111. [CrossRef]
18. Wernke, M.; Skvortsov, P.; Dürr, F.; Rothermel, K. A classification of location privacy attacks and approaches. *Pers. Ubiquit. Comput.* **2014**, *18*, 163–175. [CrossRef]
19. Alyousef, A.; Srinivasan, K.; Alrahhal, M.S.; Alshammari, M.; AL-Akhras, M. Preserving Location Privacy in the IoT against Advanced Attacks using Deep Learning. *Int. J. Adv. Comput. Sci. Appl.* **2022**, *13*, 416–427. [CrossRef]
20. Jagarlapudi, H.N.S.S.; Lim, S.; Chae, J.; Choi, G.S.; Pu, C. Drone Helps Privacy: Sky Caching Assisted k-Anonymity in Spatial Querying. *IEEE Syst. J.* **2022**, *16*, 6360–6370. [CrossRef]
21. Tadakaluru, A. Context Optimized and Spatial Aware Dummy Locations Generation Framework for Location Privacy. *J. Geovis. Spat. Anal.* **2022**, *6*, 27. [CrossRef]
22. Parmar, D.; Rao, U.P. Dummy Generation-Based Privacy Preservation for Location-Based Services. In Proceedings of the 21st International Conference on Distributed Computing and Networking (ICDCN 2020), New York, NY, USA, 4–7 January 2020.
23. Kalnis, P.; Ghinita, G.; Mouratidis, K.; Papadias, D. Preventing location-based identity inference in anonymous spatial queries. *IEEE Trans. Knowl. Data Eng.* **2007**, *19*, 1719–1733. [CrossRef]
24. Zandbergen, P.A. A comparison of address point, parcel and street geocoding techniques. *Comput. Environ. Urban Syst.* **2008**, *32*, 214–232. [CrossRef]
25. Evans, D.G.; Jones, S.M. Detecting Voronoi (area-of-influence) polygons. *Math. Geol.* **1987**, *19*, 523–537. [CrossRef]
26. Van Dongen, S.; Enright, A.J. Metric distances derived from cosine similarity and Pearson and Spearman correlations. *arXiv* **2012**, arXiv:1208.3145.

27. Zhang, C.; Liang, H.; Wang, K.; Sun, J. Personalized Trip Recommendation with Poi Availability and Uncertain Traveling Time. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, Melbourne, Australia, 18–23 October 2014; pp. 911–920. [CrossRef]
28. NYC OpenData. Department of Finance Digital Tax Map. 2022. Available online: https://data.cityofnewyork.us/Housing-Development/Department-of-Finance-Digital-Tax-Map/smk3-tmxj (accessed on 1 September 2022).
29. NYC OpenData. Points of Interest. 2022. Available online: https://data.cityofnewyork.us/City-Government/Points-Of-Interest/rxuy-2muj (accessed on 16 November 2022).
30. Esri Inc. *ArcGIS Pro, version 2.8.2*; Esri Inc: Redlands, CA, USA, 2021.
31. GIS.NY.GOV. NYS Civil Boundaries. 2022. Available online: https://gis.ny.gov/gisdata/inventories/details.cfm?DSID=927 (accessed on 1 September 2022).
32. Esri Inc. World Imagery. Available online: https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9 (accessed on 1 September 2022).
33. RStudio Team. *RStudio: Integrated Development Environment for R*; RStudio: Boston, MA, USA, 2021; Available online: http://www.rstudio.com/ (accessed on 18 December 2022).
34. Shokri, R.; Theodorakopoulos, G.; Le Boudec, J.Y.; Hubaux, J.P. Quantifying Location Privacy. In Proceedings of the 2011 IEEE Symposium on Security and Privacy, Oakland, CA, USA, 22–25 May 2011; pp. 247–262. [CrossRef]