A Semantic Obfuscation Technique for the Internet of Things

Mahmoud Elkhodr, Seyed Shahrestani and Hon Cheung
School of Computing, Engineering and Mathematics
University of Western Sydney
Sydney, Australia

***Abstract*—Although some people might willingly reveal their location information in order to obtain location-based services, few would be comfortable having their locations collected and profiled by the billions of things envisioned in the Internet of Things (IoT), at all time and in all situations. The diffusion of wireless communication networks and the technical advancements of location positioning techniques, power things of the IoT with the capabilities of automatically sensing, communicating, and processing the information about a person's location, with a high degree of spatial and temporal precision. In this work, we address the location privacy issue by introducing the Semantic Obfuscation technique (S-Obfuscation). This technique, compared to classical geometric-based obfuscation techniques, relies on geographic knowledge to produce obfuscated locations that are harder to be detected as fake or obfuscated by an adversary. The obfuscation process is supported by our novel use of ontological classification of locations based on a geographical knowledge.**

***Index Terms*—Internet of Things, Location-based Services, Privacy, Obfuscation, Ontology.**

I. Introduction

The evolved concept of communications in which ordinary and physical objects surrounding us are connected to the Internet is referred to as the Internet of Things (IoT). While the research community has different perspectives on the IoT, depending on their areas of research interests, the benefits from the outcomes of their research that will accrue to society are apparent to everyone. In fact, having various research on different areas of the IoT will ultimately lead to the realization of the vision of a world in which billions of things are communicating over the Internet, reporting their locations, identities and operational histories. The IoT enables service providers to provide services over the Internet using a completely new range of devices referred to as things or objects. These things can be connected to the Internet using wireless enabled technologies over wide area networks or within local area networks. The IoT applications range from everyday consumer devices to more sophisticated systems and networks such as smart homes, remote health monitoring systems [3], and smart grids. Beyond the massive technological opportunities and benefits of the IoT, lie important challenges. When our streets, homes, work and recreation places, and shopping centres are equipped with many connected things sending information across the Internet, recording and logging everything from a user's movements to what the user has just eaten, bought or done, unprecedented privacy challenges arise. Chief among these challenges is location privacy [4]. Recent technological advances in wireless communications, location-enabled hardware and location identification techniques will extend the location sensing capabilities to the things of the IoT. This gives rise to the possibility of using the tracking capabilities of the things for the violation of the privacy of users. Not only preserving the location privacy of users is vital, but also preserving the location privacy of the actual thing is of paramount importance. In principle, most location-based services do not require the personal identification of a user. However, even without providing any personal identification, positioning information in the form of a specific location of a thing along with the possibility of associating the location information gathered from another thing or group of things will eventually lead to revealing the movement of a thing or the user; along with other inferred personal or contextual information such as the type and nature of the activity performed. Combining all of this information with other quasi-identification information will infer other types of sensitive information such as, the thing's context, its activities, and possibly the identity of the user of the object. Our previous work in [7] described the various threats to privacy in the IoT. For instance, an automatic invasion attack is described as an incremental process of an inference attack in which the attacker gradually gathers more knowledge on the victim's life or activities through the combination and linking of the information collected from various things associated with the user on the IoT.

To confront these privacy challenges, in a previous work [6], a context-aware adaptive agent for the protection of location privacy for general devices in the IoT has been introduced. This agent employs mechanisms that adapt autonomously to different environments and activities using a proposed approach that uses a context analysis and privacy manager processes. This work introduces the Semantic Obfuscation (S-obfuscation) method as a continuation of the previous work. The S-obfuscation adjusts the granularity of location information using several levels of obfuscation. The obfuscation's levels are determined using a dynamic context analysis process, which takes into consideration the users' or things operators' privacy preferences. The agent included a mechanism for defining context disclosure policies constructed by mapping the user's privacy preferences to a particular context. The S-obfuscation method, proposed in this paper,

978-1-4799-4640-2/14/$31.00 ©2014 IEEE

augment the capabilities of the agent by generating obfuscated locations that appear to be more realistic as compared to those generated by other classical obfuscation techniques. We define a prediction rate as the ability of an adversary to detect if a received obfuscated location is real or fake. Thus we will demonstrate that the proposed S-obfuscation method generates obfuscated location with a less prediction rate. The performance of the S-obfuscation method will be evaluated against the classical obfuscation techniques, specifically the "Rand" and "Dispersion" techniques. To achieve this, an experiment using real location information is conducted using the classical obfuscation methods and the proposed S-obfuscation method. The experiment was focused on a set of location information that was randomly generated in selected geographic areas in Australia. The sample dataset of locations information was gathered by means of a GPS system. The results collected from the experiment show that the S-obfuscation method outperformed, in terms of the prediction rate, the two classical obfuscation methods under comparison.

The paper is organized as follows: Section II presents some background information and related work on the topic of location obfuscation. Section III describes the classical obfuscation techniques used in this paper for comparison purposes. Section IV introduces the S-obfuscation technique. The experimental design and implementation are reported in Section V, which also discusses the results of the performance evaluation of S-obfuscation. Conclusion remarks and future works are provided in Section VI.

II. Background

In the IoT, location information is among the most common piece of contextual data used in a wide variety of services. Examples of these services include the location-dependent information delivery services (e.g. environmental sensor readings), location-aware emergency services (e.g. general health emergency services), and location-based advertisement services. Current research on location privacy is mainly centered on two main computational protection techniques: anonymity and obfuscation. This work is concerned with obfuscation as a method to protect the location privacy of objects or users in the IoT. To perform obfuscation, three approaches have been previously proposed in the research literature. These approaches are briefly described as follows:

- Adding random noise to location information: The aim is to make it harder on the adversary to infer the actual location of the user;
- Rounding by narrowing the coarse location e.g. by using landmarks to approximate a location;
- Redefinition of possible areas of location. Example: the use of "invisible cloaking", in which no locations are provided for identified selected areas.

III. Classic Obfuscation techniques

A. The Random Technique

This technique, referred to as "Rand" in [9], is the simplest obfuscation technique.This technique alters the original location L by an another location L'. Figure 1(a) shows an example in using this technique. L' is an example of an obfuscated location generated from the original location. It works as follows:

1) For a location L, generate a circular area A with a center L and radius r. Increasing the radius r will result in enlarging the area around L.
2) Generate a random point within the circular area A e.g. L' in Fig 1(a).
3) Replace L by L'

Note that a larger r implies a larger expected noise in the generated location point L'.

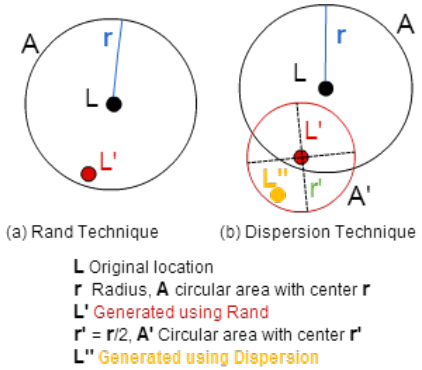

Fig. 1. The Rand and Dispersion Techniques

B. Dispersion Technique

This technique is a variant of the Rand technique. Instead of using the original location as the center of the circle, it uses a new location. Let L be the original location, r a radius to be used to generate the area A with the center L. Let L' be the obfuscated location obtained using the Rand technique. The followings is an example that uses the dispersion technique in generating an obfuscated locations:

1) For a location L, generate a circular area A with a center L and radius r.
2) Generate a random point L' within the circular area A.
3) Generate a new circular area A' with a center L' and radius r'= r/2. (The size of r' is selected randomly).
4) Generate a random point L" within the circular area A'
5) Replace the original location L with L" as shown in Fig. 1(b).

The Random and Dispersion obfuscation techniques serve as a base for several new obfuscation techniques. For instance, the N-Rand technique is an improved version of the random techniques reported above. The N-Rand introduces a new parameter that allows the selection of location points that are more distant from the original location L. Following this principle of selecting the most distant point to the original location, the dispersion technique has also been extended in such a way that L is also generated as the most distant point to the original location. In [9], these methods were compared and evaluated, and it was found that the N-Rand and N-Dispersion

techniques produced a larger minimum distance to the original location, and the greatest average distance to the original path.

C. Weaknesses in the classical obfuscation techniques

The classical location obfuscation techniques, such as the Random and Dispersion techniques, reported in the previous Section, employ geometric methods to generate obfuscated locations. These techniques are referred to as the geometric-based techniques in [1]. These techniques do not account for geographical knowledge. Thus, they do not take into consideration of the actual geographical environment at a generated obfuscated location. In real life, a location may be in a public place, a private location such as inside a building, which is closed during that time of the day, somewhere in the middle of the desert, or the ocean. Geometric-based obfuscation techniques ignore the semantics behind a geographic location. Consequently, the authors in [2] claim that an adversary with sufficient geographical knowledge may be able to infer sensitive location information from obfuscated locations generated by geometric-based techniques. Our work aligns with this claim. We will show that when a geometric-based obfuscation technique is applied to a physical environment, certain obfuscated locations are more likely to be identified as obfuscated locations by an adversary. In Section VI, we will discuss this in more details.

IV. Semantic obfuscation technique

The semantic obfuscation technique (S-obfuscation) proposed in this work uses geographical knowledge to guarantee that a generated obfuscated location is reasonable in the context of the location's environment. That is, the method avoids generating an obfuscated location that might be easily identified as obfuscated by an adversary. For example, if an obfuscated location is in the middle of a lake or sea, then it is easy for the adversary to determine that either the subject is on a boat or the location received is not a real one. Further knowledge on the geographic location, such as whether boats are allowed on that location or not, might help the adversary in identifying the received location as not original. To incorporate local geographical knowledge in the generation of obfuscated locations for an object, we propose a novel method that uses an ontological classification of geographic locations based on geographical knowledge. Our ontology is based on the one proposed in [10]. In our new approach to obfuscation, an original location L, consisting of longitude and latitude coordinates is converted into ontology as shown in Fig. 2. Generally, ontologies are used as structural frameworks for organizing information and concepts within a domain, and the relationship between those concepts. We use ontology as a form of knowledge representation about the geographic knowledge of a location. The ontology, presented in Fig.2, is part of a larger ontology that follows a top-down representation of several classes and subclasses arranged in a hierarchy as shown in Fig 3. Note that in Fig. 2, only a single path of the ontology is presented from Fig3. The ontology classes are constructed based on the naming of the geographic

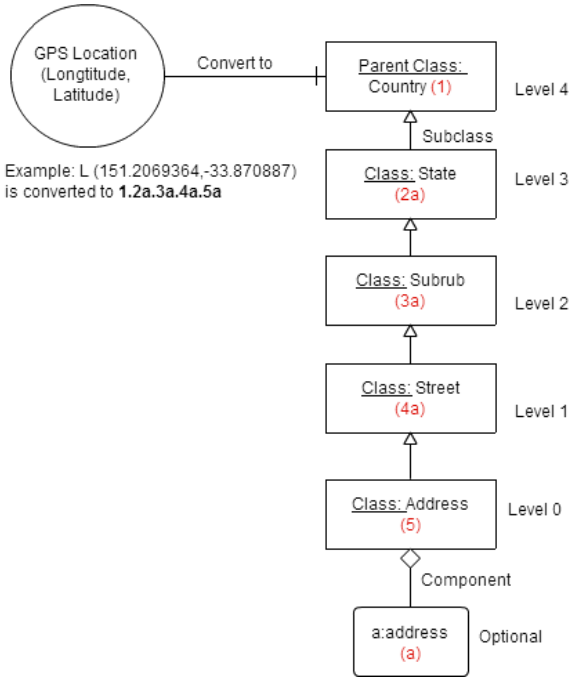

Fig. 2. The Location Ontology

subdivisions used for address purposes in Australia. Therefore, in this work, the ontology is limited to Australia.

The parent ontological class in the hierarchy has n child classes. Each of these child classes is a parent of another set of subclasses, which in turn are more general than their child classes. This top-down hierarchal distribution represents the relative proximity of the obfuscated location to the original true location. The parent class has the widest proximity in the hierarchy, while objects of the lower classes represent a smaller proximity. Thus, proximity decreases the further we go down in the hierarchy and increases the further we go up. Ontological classes on the same level of the hierarchy are considered to be of the same granularity. These classes are structured in a tree starting from the tree root country (1), where 1 is the node identifier, as shown in Fig. 2 and 3. Each class in the tree has a name, an identifier, and a node address. For example, the class "Street (4a)" has the name "Street" and the identifier "4a". The node address is constructed by the set of identifiers, separated by a dot (.). This node address is used to define the path to the class from the tree root. All classes start with the number 1. For example, the node address for the class with the name "Street" and identifier "4a" is constructed by following the path starting from "country(1)" through "state(2a)" and "suburb(3a)" down to "street(4a)". Thus the node address of class "street(4a)" is: 1.2a.3a.4a. All classes in the tree are identified using this scheme.

The S-obfuscation technique works as follows: For each object belonging to a class of Level 2 an object from a Level 1 class is selected and defined as a base point (an object is an instance of a class). The base point is chosen based on

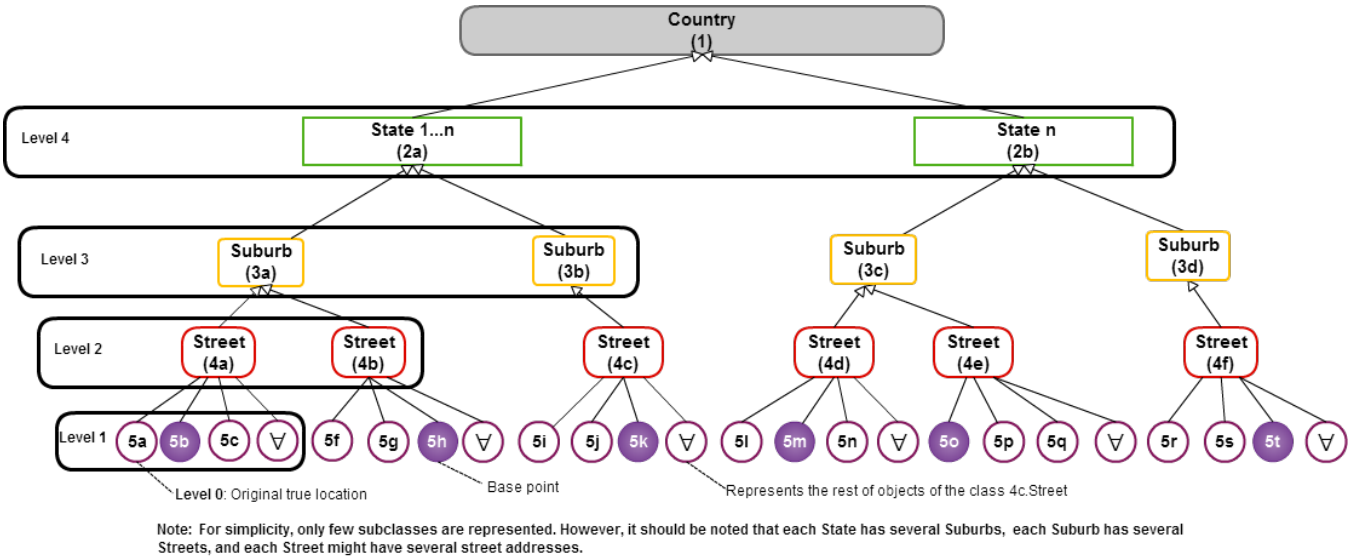

Fig. 3. The Location Tree Ontology

the geographic knowledge. For example, from Level 2, the class "Street (4a)" has subclasses with the following node addresses: 1.2a.3a.4a.5a; 1.2a.3a.4a.5b; and 1.2a.3a.4a.5c. The object with the node address 1.2a.3a.4a.5b is defined as the base point for all objects of the class "Street(4a)". Similarly, the object with the node address 1.2a.3a.4b.5h is defined as the base point for all objects of the class "Street(4b)" and so on. Therefore, any object with a node address of 1.2a.3a.4a.X (where X represents the identifier of the object of a Level 1 class) will be obfuscated to 1.2a.3a.4a.5b. Similarly, for the class "Street(4b)", any object with a node address of 1.2a.3a.4b.X will be obfuscated to the defined base point 1.2a.3a.4b.5h, and so on. Consequently, the following base points are defined for all classes of Level 2: 1.2a.3a.4a.5b, 1.2a.3a.4b.5h, 1.2a.3a.4c.5k, 1.2a.3a.4d.5m, 1.2a.3a.4e.5o, and 1.2a.3a.4f.5t. For simplicity in this example, only one base point has been defined in each class. Defining multiple base points will increase the number of available obfuscated locations that can be used.

A. *Scenario:*

Let L (lon, lat) be a true original location and L the new obfuscated location. Convert L to ontology with a node address of 1.2a.3a.4a.5a.

Suppose Bob is at a location that he considers sensitive. This location has a GPS location of L (151.2069364,-33.870887) which corresponds to the physical address 22 George Street, Sydney, NSW, Australia. Suppose that Bob would like to search for a restaurant located on George Street as he doesn't want to walk far. While Bob doesn't mind revealing that he is on George Street, he prefers not to disclose his exact location. To do that, Bob uses Level 1 of the S-obfuscation technique. Therefore using Level 1, for L with the correspondent node address 1.2a.3a.4a.5a, the obfuscated location L' with the correspondent node address of 1.2a.3a.4a.5b will be used.

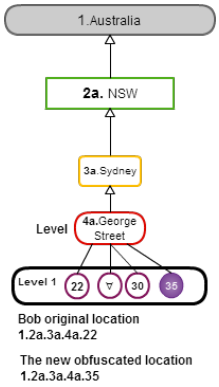

Fig. 4. Obfuscation Level 1

Fig 4 shows that instead of sending Bob's true location of 22 George Street, Sydney, NSW, Australia, an obfuscated location 35 George Street, Sydney, NSW, Australia, is used. By obfuscating his true location L with a location L' located on George Street, Bob will be able to obtain services from the LBS provider without revealing his exact location. In addition, since L' is already carefully selected based on geographic knowledge, Bob is confident enough that any adversary is unable to detect if Bob is sending a fake location. In another scenario, Bob preferences are changed to as follows: Bob does not wish to reveal his exact location, but he doesn't mind letting an adversary know that he is at somewhere in Sydney. For this, Bob uses Level 2 of the S-obfuscation techniques. Instead of sending his true location L, with the node address 1.2a.3a.4a.5a, Bob is now able to choose to send an obfuscated location, with a wider proximity to his true location than before. This Level 2 obfuscated location is selected based on geographic knowledge as well. Fig. 5 shows several options of obfuscated location that Bob can use. Level 3 of S-obfuscation

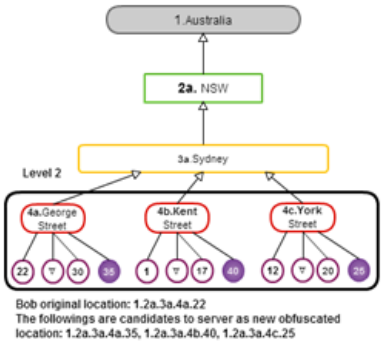

Fig. 5. Obfuscation Level 2

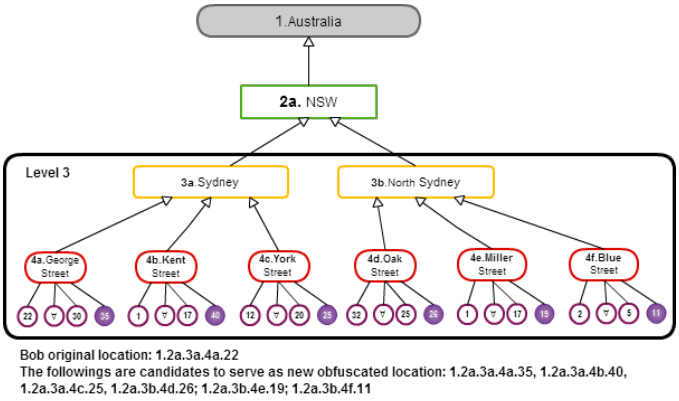

Fig. 6. Obfuscation Level 3

will allow Bob to select an obfuscated location with a wider proximity to his true location than those used in Levels 2 and 1 as shown below in Fig. 6; while Level 4 of S-obfuscation will allow bob to select an obfuscated location with a wider proximity to those generated in Levels 3, 2 and 1.

V. Evaluation

In this section, the performance of the S-obfuscation in comparison with the Rand and Dispersion obfuscation techniques is evaluated. The performance metric measured in the experiment is the prediction rate. A prediction rate is the ability of an adversary to detect if a received obfuscated location is real or fake. In order to determine if an obfuscated location is real or fake, the obfuscated location is converted using the ontology to a real address and the node address of the obfuscated location is examined. If the node address is less than 3 levels down from the root node then the obfuscated location is deemed as fake. For example, converting an obfuscated GPS location to an address, a precise address with enough geographic knowledge would look like: Australia, NSW, Sydney, George Street, and Street number (optional). If the obfuscated location does not possess enough knowledge to position a location to a street address, then the location is considered to be too ambiguous. Therefore the obfuscated location is judged as fake. This process is illustrated in Fig 7.

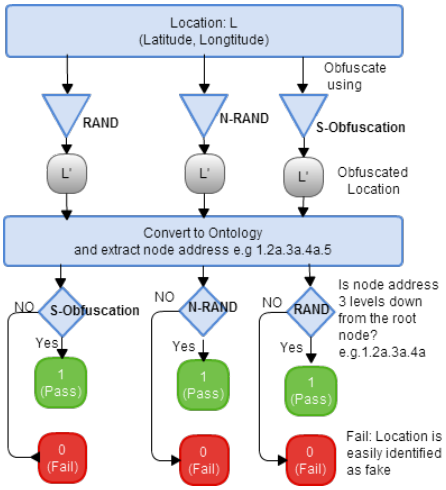

Fig. 7. Evaluation process

A. Implementing the Rand and Dispersion Techniques

The Rand and Dispersion techniques described in section III were implemented as part of a web application [5]. This application is used to test and evaluate the obfuscated location generated by these techniques. It allows the user to randomly generate obfuscated location for specific GPS coordinates. In addition, it gives the user a way of varying the radius of the area where obfuscated locations are generated. Listing 1 shows the main part of the code used to implement the Rand technique. The Dispersion technique is based on this code as well.

```
//Generate Random Numbers between -1 and 1
var u= ((Math.random()*-1)+1)*1;
var v= ((Math.random()*-1)+1)*1;
//convert the radius r into degrees,there are about
    111,300 meters in a degree.
r= Number(r) /111300;
var w = Number(r) * Math.sqrt(u);
var t = 2 * Math.PI * v;
var x = w * Math.cos(t);
var xx = x / Math.cos(y0)
var y = w * Math.sin(t);
//Generate the new location using the above
var xnew=Number(x0)+Number(xx);
var ynew=Number(y0)+ Number(y);
```

Listing 1. The Rand technique codes

B. The S-obfuscation development

The S-obfuscation is implemented in a mobile application as part of the Dynamic Location Disclosure Agent (DLDA) introduced in [6]. The application implements the different levels of obfuscation described in Fig. 3. It gives the user the ability to obfuscate his or her location to one of the base points defined in each of the S-obfuscation levels. It works as follows:

- Convert the GPS coordinates of an original location to an address using the Google Map API. Example: Convert the GPS location of L (-33.870887, 151.2069364) to 452 George Street Sydney NSW Australia

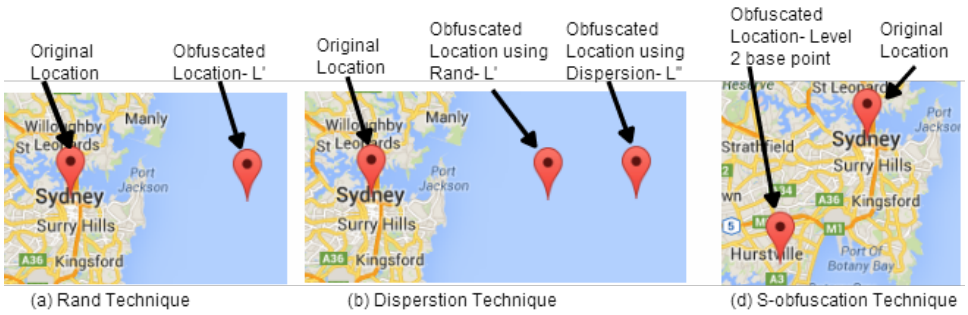

Fig. 8. Performance Example.

- Use the proposed ontology to extract the classes, their identifiers, and the node address of the converted GPS address. That is 452 George Street Sydney NSW Australia will correspond to: Australia (1), NSW (2a), Sydney (3a), George Street (4a), 452. Therefore the node address is 1.2a.3a.4a.452
- Based on the obfuscation level indicated by the user, a base point will be selected from the ontology. For instance, in the same example above, suppose the user has selected obfuscation Level 2. Therefore, following the tree structure of Fig. 3, the node address will be shortened to 1.2a.3a. From this point, a search for base points is conducted on all subclasses which start with a node address of 1.2a.3a. The search will return all base points defined below the subclasses with the identifier "3a". Next, a base point is randomly selected, e.g. 1.2a.3a.4a.5b or 1.2a.3a.4b.5h
- 4. The selected base point is then converted, using Google map API, to a GPS coordinates and used as an obfuscated location for the original location.

C. Experimental results:

In order to compare the performance of the S-obfuscation technique against the Rand and Dispersion techniques, three obfuscated locations for a given GPS original location have been generated. The three obfuscated locations are then plotted using Google map and converted to real addresses as shown in Fig 8. Next, the predication rate of each of the obfuscated location is calculated using the process described in Fig. 7. The experiment used a sample of 250 GPS coordinates randomly selected from areas within the state of New South Wales in Australia and the results were noted. The results show that the S-obfuscation technique has outperformed both the Rand and Dispersion technique in term of prediction rates.

VI. Conclusion

In this paper we presented the S-obfuscation technique, which enhances the location privacy of users. Unlike other classical obfuscation mechanisms that employ geometric methods to generate obfuscated locations, the S-obfuscation technique relies on a geographic knowledge to produce obfuscated locations that are harder to be detected as fake or obfuscated. The results of the experimental works show that the proposed technique has outperformed, in term of prediction rate, the performance of the classical Rand and Dispersion techniques. Future work will be concerned in generalizing the ontology in a way it can be adopted universally and not only in Australia. Also, future works will incorporate the S-obfuscation, as a location privacy preserving technique, in a proposed management platform for the IoT refereed to as the Internet of Things Management Platform.

References

[1] M. L. Damiani, E. Bertino, and C. Silvestri, "Protecting location privacy through semantics-aware obfuscation techniques," in *Trust Management II*. Springer, 2008, pp. 231–245.

[2] M. L. Damiani, E. Bertino, and C. Silvestri, "Protecting location privacy against spatial inferences: the probe approach," in *Proceedings of the 2nd SIGSPATIAL ACM GIS 2009 International Workshop on Security and Privacy in GIS and LBS*. ACM, 2009, pp. 32–41.

[3] M. Elkhodr, S. Shahrestani, and H. Cheung, "Enhancing the security of mobile health monitoring systems through trust negotiations," in *Local Computer Networks (LCN), 2011 IEEE 36th Conference on*, 2011, pp. 754–757.

[4] M. Elkhodr, S. Shahrestani, and H. Cheung, "A review of mobile location privacy in the internet of things," in *2012 10th International Conference on ICT and Knowledge Engineering, (ICT Knowledge Engineering)*, 2012, pp. 266–272.

[5] M. Elkhodr. (2013) Classic and dispersion obfuscation techniques implementations @ONLINE. [Online]. Available: http://elkhodr.com/obf.html

[6] M. Elkhodr, S. Shahrestani, and H. Cheung, "A contextual-adaptive location disclosure agent for general devices in the internet of things," in *The 38th IEEE Conference on Local Computer Networks (LCN)*, Sydney, Australia, Oct. 2013.

[7] M. Elkhodr, S. Shahrestani, and H. Cheung, "The internet of things: Vision and challenges," in *TENCON Spring Conference, 2013 IEEE*, 2013, pp. 218–222.

[8] L. Liu, "Privacy and location anonymization in location-based services," *SIGSPATIAL Special*, vol. 1, no. 2, pp. 15–22, 2009.

[9] P. Wightman, W. Coronell, D. Jabba, M. Jimeno, and M. Labrador, "Evaluation of location obfuscation techniques for privacy in location based information systems," in *Communications (LATINCOM), 2011 IEEE Latin-American Conference on*. IEEE, 2011, pp. 1–6.

[10] R. Wishart, K. Henricksen, and J. Indulska, "Context obfuscation for privacy via ontological descriptions," in *Location-and Context-Awareness*. Springer, 2005, pp. 276–288.