Swap Obfuscation Technique for Preserving Privacy of LBS

Yazed Alsaawy[1], Ahmad B.Alkhodre[*1] ,Adnan Ahmed Abi Sen[2] , Muhammad Shoaib Siddiqui[1]

[1]Faculty of Computer and Information Systems, Islamic University of Madinah, Al-Madinah, Saudi Arabia
[2]College of Computing and Information Technology, King Abdul-Aziz University, Jeddah, Saudi Arabia

[*]**Corresponding author:** Dr. Ahmad B. Alkhodre
E-mail: aalkhodre@iu.edu.sa

Abstract

Smart cities rely on Internet of Things (IoT) enabling technologies for ubiquitous implementations, such as, Wireless Sensor Networks (WSN), which are used to sense the environment and RFIDs, which provide object tracking of resources and assets based on identity and Global Positioning System (GPS) for localization and positioning. In smart cities, most of the health and transportation services rely on the Location Based Service (LBS), in addition to different computing models, such as, cloud and fog to work on huge amount of data collected throughout the smart city at every moment. Smart cities must enable their services anywhere and anytime to provide smarter services and with user adaptation. However, privacy has emerged as another serious challenge in addition to security as a consequence of these 'user-centric' and 'always connected' services. This paper discusses the problems of privacy and introduces a new privacy protection policy called Exchange of Confused Areas. Through simulation and testing, this research demonstrates its superiority over most of the previous approaches by providing a higher level of privacy, non-reliance on a trusted party, and improved performance.

Keywords: *Privacy, IOT, smart cities*

I. INTRODUCTION

Smart Cities promise a vibrant change in the lives of users by providing smarter and more adaptive services to users through a wide range of smart applications in various fields, including energy, sustainability, health, and transportation [1]. Because of the diversity of services, smart cities are the most important application of the Internet of Things. The number of Internet-connected devices are projected to reach more than 50 billion till 2020 [2], where WSNs in addition to RFID would make up the largest percentage of the IoT devices. Sensors are deployed in smart cities everywhere to monitor and sense information about the surrounding environment or users. The sensors include temperature, pressure, pollution, noise, light, movement and many other sensors. The RFID are used to distinguish objects and track them. The development of smart and adaptable services that can be accessed with a single click, are based on the collected data by the sensors and localization data provided by the identifier and tracking system using RFIDs [3].

As most of the IoT devices have limited resources, most applications rely on third-party services and technologies, such as, cloud computing, to store, process, analyze, and provide permanent access to information; fog computing for data pre-processing before sending to the cloud; and GPS related information, which is a key element used in most modern applications in smart cities, such as, looking for points of interest, communication, traffic regulation, rapid response to emergencies, and other technologies and applications [4-6]. The following figure shows a general perception of smart cities and some of their applications:

Fig1. Perceptions of smart cities and their applications

Unfortunately, there are challenges for the future of smart city's enabling technologies and its smart tools and sophisticated applications, which are security and privacy of users' information in these environments. The information is collected everywhere, all-the-time for every user and transmitted over the network and permanently stored in a cloud. If we assume the existence of a malicious entity that can access this information, it means we face threats to privacy of data in terms of disclosure of sensitive data and exposure of private information about the users, which might reveal a lot about the behavior of user's life, habits, ethics, relations, work and many others aspects [7].

To increase the level of security of data and information while transferring, or/and storing, from any external attacker, most of the applications rely on the use of certain cryptographic mechanisms and protocols [8]. However, what if the service provider itself is a malicious entity that can access the entire data at any time and then analyze them to discover additional information outside of the declared service, in which case we are dealing with another level of threat more difficult to encounter than the conventional one, which is called a breach of privacy. To deal with this threat, some approaches have been proposed, but unfortunately, it is still an open issue [9-10]. Some have taken the easier approach and considered that the service provider is reliable and does not pose a threat to the privacy of its users' data and while other have proposed solutions which impact on the quality of the basic services, significantly. Furthermore, they impose overhead on the service provider and effects performance in achieving high level of privacy protection, such as, in Dummy, Confusion, and Private Information Retrieval PIR [11] The issues are further discussed in the next section.

This article attempts to develop a solution for both the idea of Dummy and the idea of confusion to create a new integrated privacy protection approach that achieves the advantages and avoid disadvantages of the previous methodologies. The protection of privacy in location-based applications (as shown in Figure 2) requires protection of several parts of the user's query which is sent to the service provider such as, the sender's identity, the requested query (what is being searched for), the time of sending the query, and finally the current user location, as shown in Figure 2.

The contributions to this research article are:

- Create a new approach called Swap Obfuscation Approach to protect privacy in the location-based applications of smart cities.
- Solving the problem of dummy approach in terms of generating undetectable dummy by the service provider at the same time without increasing the load on the user.
- Solve the problem of result inaccuracy in confusion approach in addition to avoiding overloading the user during the use of large confusion in achieving acceptable level of privacy.
- Provide simulation and better results in comparison to approach proposed previously in terms of the level of privacy and performance.

The rest of the research will be as follows. The second section will present a literature survey of the previous work on privacy protection, explain the proposed approach, analyze the possible attacks and recommend the proposed approach, and finally the results, comparison, and recommendations.

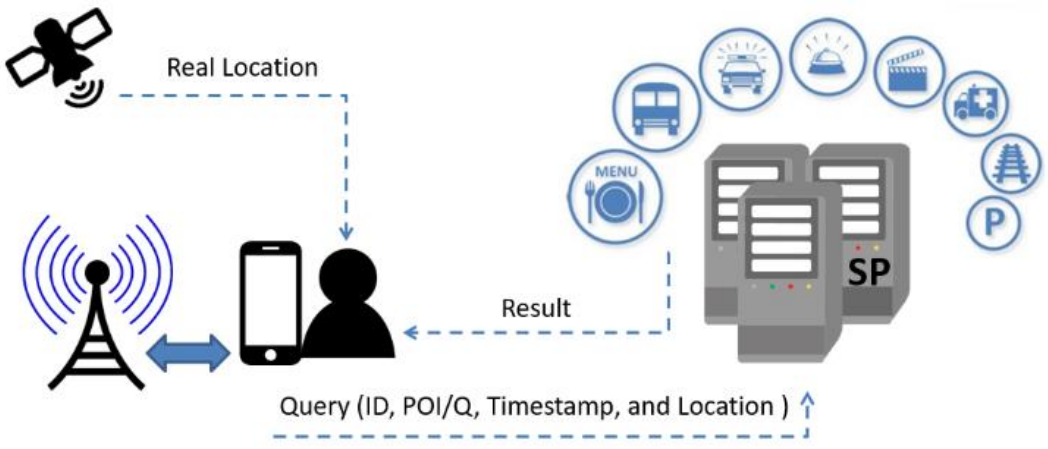


Fig 2: privacy in location-based applications

II. Previous Works

With the advancement in the level and nature of the services provided, modern technology and tools, many techniques and approaches have been proposed for advanced privacy protection in the IoT applications in various fields. Figure 3 presents fo ur classifications of protecting privacy according to the need of trust on another party [11-13].

The first classification includes methods which consider service provider as a trusted party that is committed to privacy protection laws. These methods aim at reducing the amount of collected data by the service provider, or by providing access control to the user to access his data at any time or requesting the user’s consent when the service provider wants to work on user's data. In these methods, privacy awareness is the responsibility of the user and require the user to know how to maintain data privacy by understanding service provider's policies.

The second classification involves techniques that rely on a trusted third party (another server) which maintains the privacy of the user from the service provider, such as sending a user query on his behalf, or hiding his identity within a group of users in a particular area, and/or sending a single query with the same coordinates as the trusted server zone. For increased privacy, MIX

ZONE offers the idea of changing the user name when it moves from one zone to another and when dealing with a new server to protect its privacy.

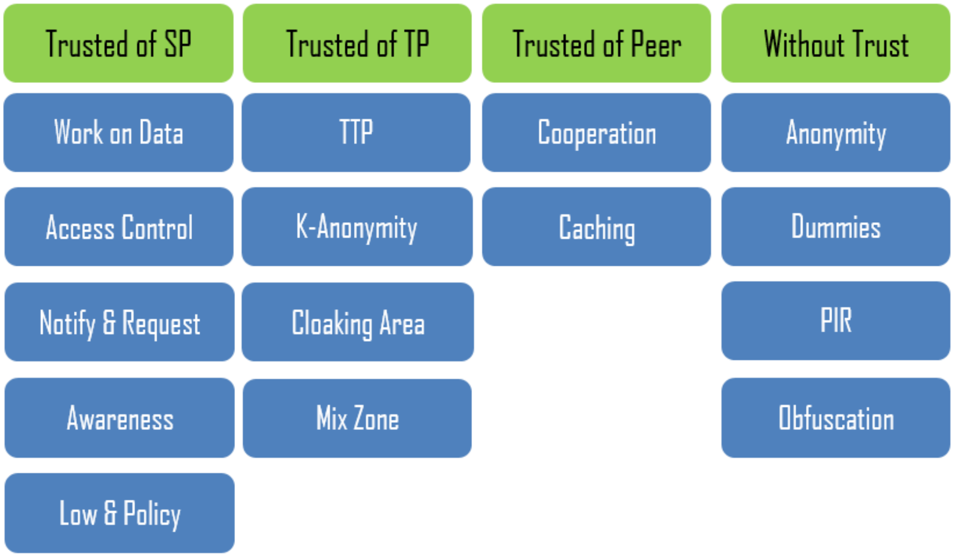
Third classification tries to avoid the need for trust in the server through the cooperation of users to protect their privacy and provides several methods of cooperation, such as relying on the cache of each user to reduce communication load or through the agreement of users to send the same location or the same query at the same time.

The fourth classification provides ways to protect privacy without having to trust any party by using an alias, sending dummy with the real query to the service provider, or requesting a large amount of data from the service provider so that it could not actually know what the user wanted or add noise to the data before sending it to the service provider.

Unfortunately, despite the existence of these classifications and techniques, there are still problems and issues. The first classification suffers from confidence in the service provider, which is considered the most dangerous to privacy in the case of malicious service providers. The second classification moves the problem faced in first from the service provider to a third-party server. In the third classification, there is an issue of trust in all users, and finally the fourth classification is causing the impact on the level of performance and significantly increases the load on the user and finally affects the accuracy of the results returned by the service provider, which affects the accuracy of the service itself [12].

This research article focuses on the last classification considering the provided technology can fall under the same classification and considers the solution to the problems in the techniques of this section, which are:

- Ineffective concealment of identity/alias in protecting the privacy of the user from the service provider [14].
- The effect of the use of dummy on the performance of the system and overhead on the user, in addition to delay in detection of false information in comparison to authorized data, and degradation in accuracy of results in some applications [15-16].
- High computational load that a user must be able to handle as well as the server, which is unrealistic in PIR [17]
- The results are significantly affected when overcrowding is used to achieve better privacy, which increases the load on user in processing the messages [18].



Trusted of SP	Trusted of TP	Trusted of Peer	Without Trust
Work on Data	TTP	Cooperation	Anonymity
Access Control	K-Anonymity	Caching	Dummies
Notify & Request	Cloaking Area		PIR
Awareness	Mix Zone		Obfuscation
Low & Policy			

Fig 3: Classification for privacy approaches Research Methodology

III. Proposed approach

This research article proposes a new approach resulting from the integration of several previous technologies (user collaboration, Dummy, confusion) in a new way of collaboration that provides new features and a better privacy level, as well as, a good solution to most of the challenges or problems of previous methods.

The main idea in this way is the SWAP; where two peer users will collaborate to increase the level of protection for both (mutual benefit).

To clarify the idea, suppose that A wants to send a query to a Service Provider (SP). At the same time A wants to protect its privacy from the same service provider and therefore A will follow the protocol:

- A will search for another random user B, who is using the same proposed approach within the same zone or in any other zone.
- A will send its query to B rather than SP.
- Since A does not trust B, it will also add a slight blur to its query before sending it to B
- B receives the query from A but cannot locate A because of confusion and because it does not have a history record for previous A queries to break this confusion (if B is also malicious)
- B will forward the query to the SP on behalf of A. B will give misleading information about itself to SP which means increasing the privacy level of A
- The service provider will not be able to collect any information about A as it has recorded the wrong information it has about B
- The service provider will reply the query and return it to B
- B returns the result to A

If another query is received, A will repeat the previous steps but with another random user and so on

.

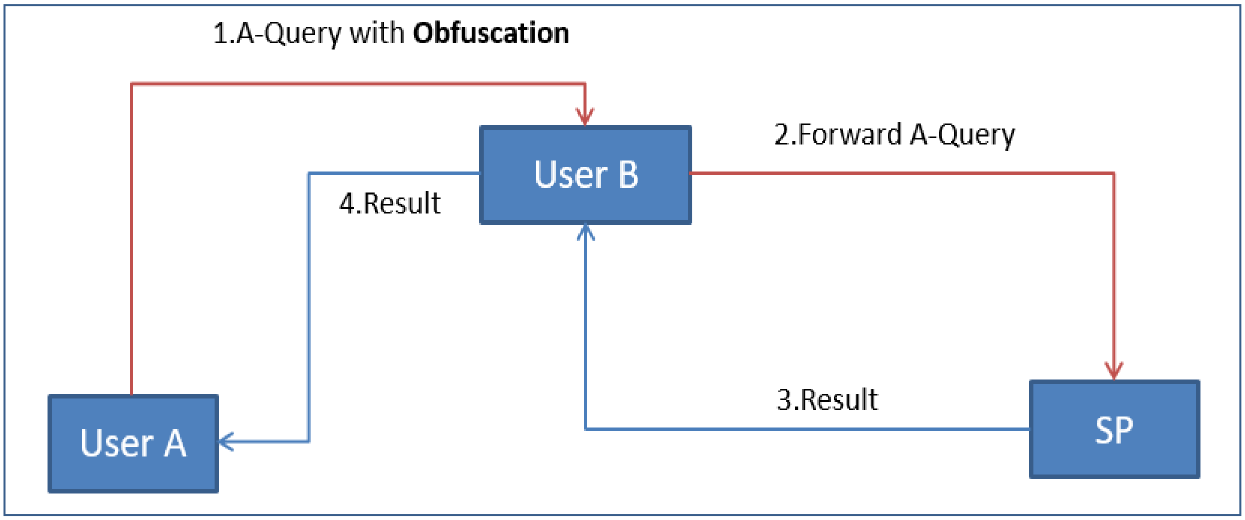


Fig 4: Proposed scenario to increase the level of privacy

To increase the level of privacy further, the proposed approach assumes that A can change its alias to communicate with another user, which increases the level of privacy very significantly even if A is dealt with B more than once during different periods. Figure 4 illustrates the proposed scenario.

a. The positives of the proposed approach

- The proposed approach provides protection for user A's privacy almost entirely from the service provider, which is considered the most dangerous if it is malicious as the user is not communicating with the service provider permanently during its queries.
- The proposed approach provides protection to the other collaborator B so that it provides a new and easy mechanism for generating a real smart dummy that cannot be detected by the service provider (which is an open issue in the dummy approach) and at the same time misleading information about sender B (user Collaborator B).
- Achieving mutual benefit between the cooperating peer and thus encouraging everyone to cooperate and without overloading any of them.
- The suggested approach always sends the service provider one query instead of a set of queries as in dummy, so the proposed approach significantly improves performance compared to the dummy approach.
- The use of simple confusion and alias switching during the process of cooperation between peer means protecting the privacy of collaborators from each other and it was also considered an open issue within the approach of cooperation.
- There is no need to use large confusion here, since the collaborator B cannot record a historical record of User A queries, so the accuracy of the results will be only slightly affected, which means solving the problem of the confusion approach, which is also an open issue so far.

In a nutshell, the proposed approach improves the level of privacy of all users (peer) significantly and increase against the service provider (malicious) with each new query sent to SP. While at the same time, significantly improving the level of performance and accuracy of results and the level of privacy among the peer themselves as well. Which is a solution to many of the open problems in the previous approaches and thus a marked advantage of the proposed approaches. Which will be demonstrated within the results section.

b. The disadvantages of the proposed approach

Despite the discussed positives in the proposed approach, it still suffers from some negatives but are not significant:

- The use of noise, even slightly, will have a slight impact on the accuracy of the results, which may not be acceptable in some applications, such as medical.
- If peer B delays returning the result to A, then A will have to re-send the query to another C, which may affect performance slightly, especially with the issue of mutual benefit

In the case of cooperation between B and SP, the privacy of A can be broken by the service provider, but it is illogical and requires the cooperation of many peers with malicious SP.

IV. Attacks that could be addressed by the proposed approach

Semantic Context: If the attacker has already personal information about the peer (like his job), it can use this data to disclose the user's queries. In the Swap Obfuscation Approach (SOA), the user does not send any real query to SP. Moreover, when user B sends a query of another peer to SP, B will mislead the SP [11- 13].

Path Tracking: This attack can be used in the traditional Dummy Approach, where the SP can connect between dummies to find the real path of users after a few times. However, in SOA, each time the peer sends a query to another random user, so SP can’t find any path from the collected data about any peer [11- 13].

Historical Data: Attacker (Malicious SP) collects a lot of data/queries for each peer/user over time, then analyzes it to detect new and more information about the user. However, in SOA, more collected information means more misleading information for SP.

Inversion Attack: If the attacker discovers the algorithm protection of the used technique, it will be able to break this protection. However, in the SOA that will not affect because the user doesn't send anything related to it, so SP can just expect that this data/query doesn't belong to the sender [11- 13].

Knowledge about the map: If the attacker has this skill, it tries to remove many of the illogical dummies, and has more information about the real one. However, in the SOA, all sent queries are real queries for real users, so SP cannot eliminate any queries [11-13].

Malicious Peer: This open issue exists in the cooperation approach; however, in SOA, the small obfuscation will solve this issue, especially the user must change its alias each time and then select random peer for dealing with it [11- 13].

V. Results and analysis

The main metric for comparing privacy approaches are the level of privacy, in addition to the performance [11, 12, 13, 19, and 20]. Each metric has sub metric, which are:

A. Privacy level [19-20]

The privacy level can be measured by the amount of correct information gathering by the attacker (which can be the SP) for each user. So, it can be determined by two basic parameters:

K-Anonymity is the number of real queries compared to the dummies queries collected by attacker about his victim.

$$K - Anonymity = \frac{1}{k} \qquad (1)$$

Where, K is the number of queries that sent to SP in each sending operation

Entropy refers to the amount of information which the attacker can link to a specific user (possibilities). Therefore, it is limited between 0 and 1. In the privacy domain, the higher value means a higher degree of privacy.

$$E = \sum pi * \log_2(pi) \qquad (2)$$

Where, p_i is the probability that q_i is related to a specific user. Other factors in privacy issue are the **Ubiquity** and **Estimation Error**, which are derivatives of the entropy metric.

B. Performance and Cost [19-20]

There are several metrics which are related to performance:

The Number of queries that are sent to the service provider in each individual operation.

Size/amount of replies, which will be returned by the SP to a user.

Accuracy of results and the time that peer needs to map these results to his real location.

Cache Hit Ratio if the metric used the caching technique.

Comparison

We have provided comparison between SOA and the traditional dummies approach, in addition to a comparison between SOA and the Enhanced Cache approach that uses caching with dummies to reduce the number of connections with the server [19-20]. We assumed that the hit rate is 35%, K = 5, and each user will send one time.

SOA shows superiority on other approaches in terms of privacy level and performance. That is because, SOA always sends only one query to SP instead of a set of queries and in the same time the sent query is not related to the sender, so SP will collect misleading information about users cumulatively with each new query received. Moreover, the SOA presents a new method to generate dummies by swapping among peers without the need to trust among them.

Figure 5 and 6 depict the results of comparisons, where Figure 5 shows the number of queries according to the number of users in the system. Figure 6 shows the amount of right information that SP can detect with each a new query.

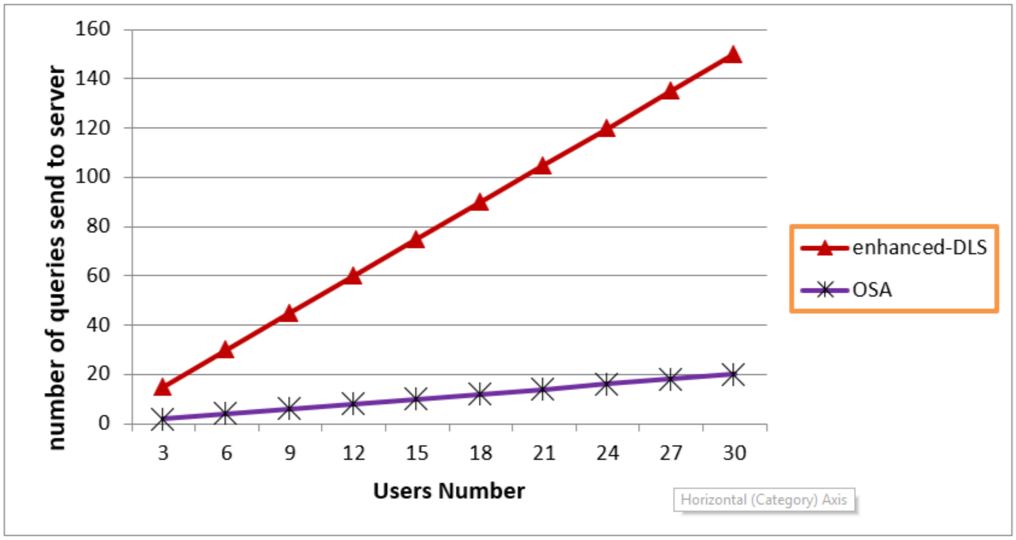


Fig. 5 Number of sent queries as compared to the number of users

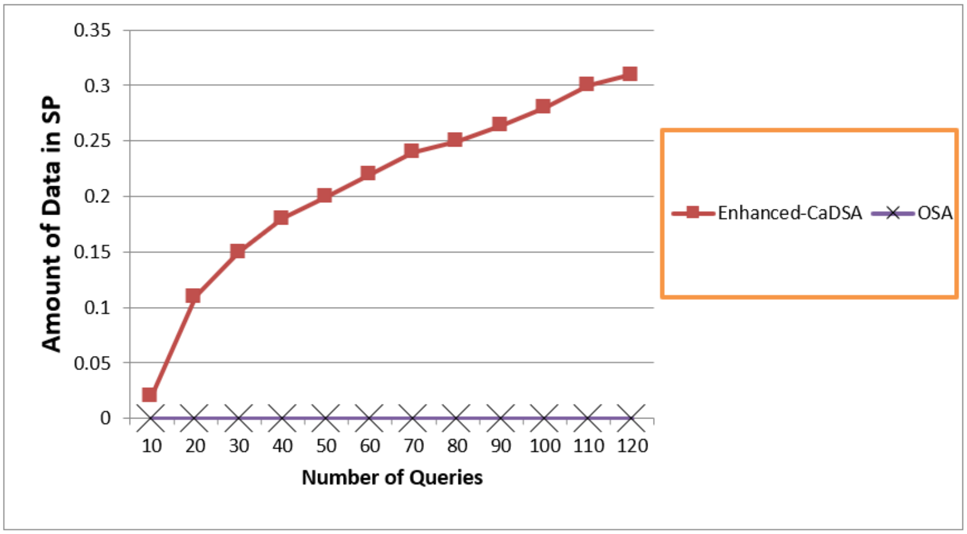


Fig. 6 Amount of correct information generated by the service provider with increases the number of queries.

VI. Conclusion

This research article proposed a new method Swap Obfuscation Approach (SOA) for preserving the privacy of IoT application especially LBS. This technique enhanced the privacy more than previous approaches and overcame the previous approaches in terms of privacy and performance. In addition, SOA solved some of the open problems in the previous approach which are (generating strong dummies an able to be detected by SP, trust issue between peers, the large adversely effects on the accuracy of the result in traditional obfuscation approach, and the performance issue in the previous approaches). Comparison with previous approaches supports our claim. In future, we will try to provide a novel approach or technique for preserving privacy in any IoT application not only LBS.

REFERENCES

[1] Roshan, R., Sharma, A., & Rishi, O. P. (2019). IoT Platform for Smart City: A Global Survey. In Emerging Trends in Expert Applications and Security (pp. 197-202). Springer, Singapore.

[2] Ammar, M., Russello, G., & Crispo, B. (2018). Internet of Things: A survey on the security of IoT frameworks. Journal of Information Security and Applications, 38, 8-27.

[3] Yamin, M., Basahel, A. A., & Abi Sen, A. A. (2018). Managing Crowds with Wireless and Mobile Technologies. Wireless Communications and Mobile Computing, 2018.

[4] Fouz, F., & Sen, A. A. (2016). Performance and Scheduling Of HPC Applications In Cloud. Journal of Theoretical & Applied Information Technology, 85(3).

[5] Sen, A. A. A, Eassa, F. A., & Jambi, K. Survey of FOG Computing Properties, Roles, and Challenges (INDIACom). 2019.

[6] Bartering Method for Improving Privacy of LBS International Journal of Computer Science and Network Security Vol. 19 No. 2 pp. 207-213

[7] Sen A, Albouraey F, Jambi KA (2017) Preserving privacy of smart cities based on the fog computing. In: Smart societies infrastructure, technologies, and applications (SCITA), Springer

[8] Al-Rahal, M. S., ABI SEN, A. D. N. A. N., & Basuhil, A. A. (2016). HIGH LEVEL SECURITY BASED STEGANORAPHY IN IMAGE AND AUDIO FILES. Journal of Theoretical & Applied Information Technology, 87(1).

[9] Abomhara M, Køien GM (2014) Security and privacy in the internet of things: current status and open issues. In: Privacy and security in mobile systems (PRISMS), international conference. IEEE, pp 1–8

[10] Sen, A. A. A, & Bashal, M. A. (2019). Comparative Study between Security and Privacy. INDIACom.

[11] Sen, A. A. A., Eassa, F. A., Jambi, K., & Yamin, M. (2018). Preserving privacy in internet of things: a survey. International Journal of Information Technology, 10(2), 189-200.

[12] Shin, K. G., Ju, X., Chen, Z., & Hu, X. (2012). Privacy protection for users of location-based services. IEEE Wireless Communications, 19(1).

[13] Wernke M, Skvortsov P, Durr F, Rothermel K (2014) A classification of location privacy attacks and approaches. Pers Ubiquit Comput 18(1):163–175

[14] Niu B, Li Q, Zhu X, Cao G, Li H (2014) Achieving k-anonymity in privacy-aware location-based services. In: INFOCOM, 2014 proceedings IEEE. IEEE, pp 754–762

[15] Alrahhal, M. S., Ashraf, M. U., Abesen, A., & Arif, S. (2017). AES-Route Server Model for Location based Services in Road Networks. International Journal Of Advanced Computer Science And Applications, 8(8), 361-368.

[16] Yamin M, Sen AAA (2018) Improving privacy and security of user data in location based services. Int J Ambient Comput Intell (IJACI) 9(1):19–42

[17] Liu, S., Liu, A., Yan, Z., & Feng, W. (2019). Efficient LBS queries with mutual privacy preservation in IoV. Vehicular Communications.

[18] Saxena, A. S., Bera, D., & Goyal, V. (2019). Modeling location obfuscation for continuous query. Journal of information security and applications, 44, 130-143.

[19] Niu B, Li Q, Zhu X, Cao G, Li H (2015) Enhancing privacy through caching in location-based services. In: Computer communications (INFOCOM), 2015 conference. IEEE, pp 1017–1025

[20] Sen, A. A. A., Eassa, F. B., Yamin, M., & Jambi, K. (2018). Double Cache Approach with Wireless Technology for Preserving User Privacy. Wireless Communications and Mobile Computing, 2018.