Revisiting a Privacy-Preserving Location-based Service Protocol using Edge Computing

Santosh Kumar Upadhyaya
santosh.upadhyaya@iiitb.ac.in
IIIT, Bangalore
IN

Srinivas Vivek
srinivas.vivek@iiitb.ac.in
IIIT, Bangalore
IN

ABSTRACT

Location-based services are getting more popular day by day. Finding nearby stores, proximity-based marketing, on-road service assistance, etc., are some of the services that use location-based services. In location-based services, user information like user identity, user query, and location must be protected. Ma et al. (INFOCOM-BigSecurity 2019) proposed a privacy-preserving location-based service using Somewhat Homomorphic Encryption (SHE). Their protocol uses edge nodes that compute on SHE encrypted location data and determines the k-nearest points of interest contained in the Location-based Server (LBS) without revealing the original user coordinates to LBS, hence, ensuring privacy of users locations.

In this work, we show that the above protocol by Ma et al. has a critical flaw. In particular, we show that their secure comparison protocol has a correctness issue in that it will not lead to correct comparison. A major consequence of this flaw is that straightforward approaches to fix this issue will make their protocol insecure. Namely, the LBS will be able to recover the actual locations of the users in each and every query.

CCS CONCEPTS
• Security and privacy → Privacy-preserving protocols.

KEYWORDS
Location-based service, privacy, edge computing, somewhat homomorphic encryption, cryptanalysis

ACM Reference Format:
Santosh Kumar Upadhyaya and Srinivas Vivek. 2022. Revisiting a Privacy-Preserving Location-based Service Protocol using Edge Computing. In The 17th International Conference on Availability, Reliability and Security (ARES 2022), August 23–26, 2022, Vienna, Austria. ACM, New York, NY, USA, 5 pages.
https://doi.org/10.1145/3538969.3544432

1 INTRODUCTION

Location-based service is meant to provide real-time information based on the current location of a user by combining multiple entities like the global positioning systems, information and communication systems, and the Internet [7, 13]. User identity, location and query information are sensitive and personal, and, hence, must be protected [6, 14]. This information needs to be protected as this could potentially be misused [15]. Hence, it is in the interest of a Location-based Service (LBS) provider to protect the private information of the users to maintain its reputation, and, hence, its business itself. Ma et al. proposed a privacy-preserving location-based service using Somewhat Homomorphic Encryption (SHE) [10]. The user provides his/her encrypted location information and the encrypted query to an Edge Node (EN). The location coordinates are encrypted using an SHE, while the query is encrypted using a traditional encryption scheme. When the encrypted service request and the encrypted user location coordinates reach EN, it generates an encrypted virtual location using a standard K-anonymity technique, in turn referring to the historical location information [17]. The LBS server contains the location coordinates of many points of interest. This information is not encrypted. Depending on the user’s query, it will select a subset of these points. But this selection must be done using the encrypted coordinates of the user virtual location computed by EN and the points of interest stored as plaintexts in LBS. The main purpose is to securely choose, say, k, nearest points of interest around the user’s location [18].

The metric used here is the Euclidean distance. So the crux of the protocol of Ma et al. is an efficient privacy-preserving distance comparison protocol that is executed between EN and LBS. The detailed steps are recalled in the Section 2.

Our Contribution. We show, in Section 3, that the privacy-preserving distance comparison protocol of Ma et al., that is eventually used in determining the nearest points of interest, suffers from a correctness flaw. Namely, the output of this comparison protocol is not necessarily correct. A major consequence of this flaw is that a straightforward approach to fix this flaw would be to give out the LBS the (signed) differences of the distances. We show, for the sake of completeness, that using these differences an LBS will be able to recover the actual location coordinates in each and every user query. We also consider another straightforward modification of the protocol whereby the differences of distances are masked by an independently chosen random value but that still allows for efficient comparison. We again show that this approach too fails in preserving the privacy of the user locations. Our work demonstrates that fixing the protocol of Ma et al. is non-trivial without incurring a significant cost.

2 RECAP OF THE PROTOCOL FROM MA ET AL.

In this section, we briefly recollect the steps of the protocol from [10]. There are four different entities in the protocol:

• User
2.1 Initialization
During the initialization step, the user registration process is executed. When a user requests for a location service, CA sets up the required private and public keys for the user, EN, and LBS. The CA generates 3 pairs of public and private keys.

- \( pk_p, sk_p;pk_p \) is sent to the user to encrypt the user request. \( sk_p \) is sent to the LBS to decrypt the user query.
- \( pk_e \) is sent to the LBS to encrypt the query response using SHE [4]. \( sk_e \) is sent to the user to decrypt the same.
- \( pk_a, sk_a; pk_a \) is sent to the LBS and \( sk_a \) is sent to the user.

Note that in this protocol the FV SHE scheme is used to encrypt users’ location information. Recall that if \( m \) and \( m' \) are plaintexts and their corresponding ciphertexts are \( c \) and \( c' \), then using an SHE scheme, the encryption of \( m + m' \) or \( m \cdot m' \) can be derived from \( c \) and \( c' \) without the need to decrypt the ciphertexts. If \( \{x\} \) is a SHE ciphertext of a plaintext \( x \), then \( [m + m'] \) can be computed as \( [m] + [m'] \). Similarly, \( [m \cdot m'] \) can be computed as \( [m] \cdot [m'] \).

2.2 User Query: USER to EN
To preserve the privacy of the user location, a K-anonymity based technique is used [12]. When the user creates a query, the query is encrypted with \( pk_p \) and the encrypted user query is sent to the nearest EN. The user sends the SHE encrypted location coordinates \( [X_o] \) and \( [Y_o] \) to EN. The EN is equipped with better storage and computing power in comparison to the user device [9]. The EN only knows that the user is within its coverage area. The EN calculates the virtual address of the user by fetching the historical location information. If the current user is considered as the \( n \)th user, then EN fetches the (encrypted) location coordinates about the previous \( i - 1 \) users:

\[
X_{oa} = \frac{1}{i} \left( [X_o] + [X_{o-1}] + \cdots + [X_{o-t+1}] \right) \tag{1}
\]

\[
Y_{oa} = \frac{1}{i} \left( [Y_o] + [Y_{o-1}] + \cdots + [Y_{o-t+1}] \right) \tag{2}
\]

\((X_{oa}, Y_{oa})\) is the computed virtual location of the user. This is the classical moving average technique used in statistics [16]. \( [X_o], [Y_o] \) denotes its SHE ciphertext. The question that arises is how to compute the encryption of \( \frac{1}{i} \). One could possibly use the fixed-point encoding scheme from [1] for this purpose.

2.3 EN to LBS
The EN relays the encrypted user query to LBS. After receiving the user query from EN, LBS decrypts it using \( sk_p \) and obtains the user query as plaintext. The LBS does not have any information about the virtual address \( [X_{oa}], [Y_{oa}] \) as it is encrypted. It needs to further interact with EN to build the query response. The EN’s territorial information is available with LBS. It fetches location coordinates of the services whose information it has, and sends them to EN after encrypting these location coordinates using the SHE scheme. If there are \( n \) services supported by EN, then let the location coordinates for these \( n \) services be \( \{(x_i, y_i) : i = 1, \ldots, n\} \). The encryption of these coordinates using key \( pk_a \) are \( \{(\|x_i\|, \|y_i\|) : i = 1, \ldots, n\} \).

2.4 LBS to EN
The LBS sends \( \{(\|x_i\|, \|y_i\|) : i = 1, \ldots, n\} \) to EN. Once EN receives the information from LBS, it calculates \( \{d_i : i = 1, \ldots, n\} \), which are the squared Euclidean distances of the user virtual location to the \( n \) services, as

\[
d_i = (\|X_{oa}\| - \|x_i\|)^2 + (\|Y_{oa}\| - \|y_i\|)^2 \tag{3}
\]

In order to ensure the privacy of users locations, \( [X_{oa}] \) and \( [Y_{oa}] \) should not be sent to LBS, and so is the case with \( \{d_i : i = 1, \ldots, n\} \). Next, the LBS must somehow securely sort the encrypted (squared Euclidean) distances \( \{d_i : i = 1, \ldots, n\} \) to determine the nearest distance(s). An obvious way of sorting is to compare every pair of (encrypted) distances, and this is what is done next. From the list of \( \{d_i : i = 1, \ldots, n\} \), pick any two elements, say, \( d_{a} \) and \( d_{b} \). Let \( m \) be the maximum distance covered by EN. If the range is considered as a circular area, then \( m \) is the diameter of the circle. In this case, \( 0 \leq d_{a}, d_{b} \leq m \). EN selects a number \( l \) such that

\[
l \geq m \tag{4}
\]

The EN computes

\[
[z] = [z'] + d_{a} - d_{b} = [z'] + [d_{a}] - [d_{b}] \tag{5}
\]

\([z]\) is the SHE ciphertext of \( z \), and \( z \) is an \( l + 1 \)-bit integer whose Most Significant Bit (MSB), \( z_{l} \), depends on the value of \( d_{a} \) and \( d_{b} \). If the MSB of \( z \) is 0, then \( d_{a} < d_{b} \). Otherwise, \( d_{a} \geq d_{b} \). In order to indirectly send the value \( [z] \) to LBS, EN creates a uniform random number \( \rho \) of size \( k + l + 1 \) bits. Here, \( k \) is the security parameter. The sum of \([z]\) and \([\rho]\) is computed, and is then sent to LBS:

\[
[w] = [z'] + [\rho] \tag{6}
\]

2.5 EN to LBS
Once LBS receives \([w]\), it decrypts \([w]\) and obtains \( w \). From \( w \), it calculates \( \hat{w} \) as follows:

\[
\hat{w} = w \pmod{l} \tag{7}
\]

and then computes its SHE ciphertext \([\hat{w}]\).

2.6 LBS to EN
Let

\[
\tilde{\rho} = \rho \pmod{l} \tag{8}
\]

Note that \( \tilde{\rho} \) is available with LBS, and \( \tilde{\rho} \) is available with EN. Let \( (\hat{w_{t-1}}, \ldots, \hat{w_{0}}) \) and \( (\tilde{\rho_{t-1}}, \ldots, \tilde{\rho_{0}}) \) be the bits of \( \hat{w} \) and \( \tilde{\rho} \), respectively. The LBS encrypts each bit of \( \hat{w} \) and is sent to EN. It is proposed to compare \( \hat{w} \) with \( \rho \) and determine the MSB of \( z \), using which we can in turn compare \( d_{a} \) and \( d_{b} \). The DGK scheme [2, 3] is used for this step. LBS server then runs the DGK key generation algorithm to generate the public and private key pair. The public key is sent to EN. The below steps are run multiple times so that eventually the LBS learns the sorted order of \( \{d_i : i = 1, \ldots, n\} \). After this, the LBS builds the response to the user query and then sends it to the user.
Finally, EN would have \((\[x_{[\text{va}]}, [y_{\text{va}}]\])\).

\[ (\[d_i\] : i = 1, \ldots, n) \]

EN to LBS

EN sends \((\[\tilde{c}_{l-1}\], \ldots, [\tilde{c}_0]\)) to LBS. After LBS receives this information, it decrypts this and checks the presence of 0. If 0 is present, that, the authors claim, indicates \(\tilde{w} > \tilde{\rho}\), otherwise, \(\tilde{w} \leq \tilde{\rho}\).

The EN and LBS need to run these steps \(n(n-1)/2\) times. Finally, LBS will obtain the sorted order of \((d_i : i = 1, \ldots, n)\) without knowing anything about the values \((d_i : i = 1, \ldots, n)\). After the service locations to be sent is securely determined, the user query response is created and sent to EN after encryption with key \(pk_r\). Finally, EN relays this query response to the user, and then the user decrypts it using its secret key \(sk_r\).

3 CORRECTNESS FLAW AND SECURITY IMPLICATIONS ON THE PROTOCOL OF MA ET AL.

In this section, we point out a critical flaw in the protocol from [10] recalled in the previous section. This flaw corresponds to the steps of the protocol described in Sections 2.6 and 2.7. Recall that the idea behind these steps is to use \(\tilde{\rho}\) and \(\tilde{w}\) to determine \(z_i\), the MSB of \(z\) (see Equations (6), (7) (8)). Recall that this bit \(z_i\) is used to compare distances \(d_a\) and \(d_b\). The following is an elementary fact from arithmetic:

**Fact 1.** \(z_i\) is independent of \(\tilde{w}\) and \(\tilde{\rho}\).

This is because \(\tilde{w}\) is determined only by \(\tilde{\rho}\) and \(\tilde{z}\), and the latter is completely independent of \(z_i\).

**Corollary 3.1.** The comparison protocol from [10] does not correctly determine the comparison between encrypted distances.

The following toy examples illustrate the above observation.

**Example 1**
\[ z = 3 = (11)_2 \]
\[ l = 2 = (10)_2 \]
\[ \rho = 31 = (11111)_2 \]
\[ w = z + \rho = 34 = (100010)_2 \]
\[ \tilde{\rho} = 31 \pmod{4} = 3 = (11)_2 \]
\[ \tilde{w} = w \pmod{2^l} = 34 \pmod{4} = 2 = (10)_2 \]

**Example 2**
\[ z = 7 = (111)_2 \]
\[ l = 2 = (10)_2 \]
\[ \rho = 31 = (11111)_2 \]
\[ w = z + \rho = 38 = (100010)_2 \]
\[ \tilde{\rho} = 31 \pmod{4} = 3 = (11)_2 \]
\[ \tilde{w} = w \pmod{2^l} = 38 \pmod{4} = 2 = (10)_2 \]

In both the examples, the values of \(\tilde{w}\) and \(\tilde{\rho}\) remain the same, but \(z_i\) takes both 0 and 1.

3.1 Security Implications

The privacy-preserving comparison protocol discussed above was proposed in [10] in order to leak to LBS only \(z_i\), i.e., the result of comparison between any pair of distances. This was done because leaking the full value of \(z\) would enable an adversary to determine the original user locations. For completeness, we briefly recollect...
next the steps to recover \((X_a, Y_a)\), the original user location coordinates, when LBS obtains \([z]\). Note that since the secure comparison protocol is flawed, giving out \(z\) is a straightforward, but insecure, way of fixing the protocol that can still retain the efficiency of the original protocol. Note that fully homomorphic sorting is currently impractical to be deployed on a large scale [5].

Once the LBS receives \([z]\), it can decrypt it to obtain \(z\), and then subtract \(z^2\) from \(z\) to obtain the signed difference \(d_i - d_j, 1 \leq i < j \leq n\). This can be repeated for every pair of distances. We then end up with \(n(n - 1)/2\) equations in the \(n\) many \(d_i\) \((1 \leq i \leq n)\). Hence, LBS will be able to solve for all the \(d_i\) from this overdetermined system of linear equations.

Once, say, \(d_1\), is obtained. Then, the LBS can try to solve for \((X_a, Y_a)\) from the following equation:

\[
(X_a - x_1)^2 + (Y_a - y_1)^2 = d_1,
\]

The above equation corresponds to a circle and there can be infinitely many solutions. Note that LBS knows \((x_1, y_1)\), i.e., as plaintexts. If \((X_a, Y_a)\) are encoded as (scaled) integers, then it will only have a "couple" of solutions on an average [8]. But to keep things simple, we can write similar equations for \(d_2, d_3, \ldots\). Since three circles are likely to intersect at a single point, the LBS will very likely be able to recover the user virtual location \((X_a, Y_a)\). If there are more than two points at which these three circles intersect, then we can continue this process until we narrow down to a single point.

Once the LBS obtains \((X_a, Y_a)\), then it will try to recover the original user location coordinates \((X, Y)\). It is not unreasonable to assume that the LBS would have tried to recover the user location coordinates from the very beginning. In this case, the LBS would also know the historical location coordinates \((X_{n-1}, Y_{n-1}), \ldots, (X_{n-t+1}, Y_{n-t+1})\) used in Equations (1) and (2). Also, the value of \(t\) is typically known to LBS as part of the protocol, or else, it can be guessed as it is usually small. Then, from Equations (1) and (2),

\[
X = t \cdot X_a - (X_{n-1} + X_{n-2} + \ldots + X_{n-t+1}),
\]

\[
Y = t \cdot Y_a - (Y_{n-1} + Y_{n-2} + \ldots + Y_{n-t+1}).
\]

Here, we are assuming that the virtual location information is only computed with the actual location data. Else, what if initially the parameters \((X_{n-1}, Y_{n-1}), \ldots, (X_{n-t+1}, Y_{n-t+1})\), were randomly chosen? After \(t\) instances of the protocol have been evoked, the initially chosen random values will no longer affect the computation of \((X_a, Y_a)\). While the convergence and divergence of these moving averages is well-studied in statistics, we do not know of how to recover the individual data points, if at all it is possible. In this case, we can only recover the virtual location coordinates.

### 3.2 Another Failed Attempt

Next, we look at another straightforward method to fix the comparison protocol of [10]. The (signed) difference of distance is now masked by a random and independently chosen value \(R\). Note that \(R\) could be a possibly large value chosen independently for every difference. We then have

\[
[z] = ([d_a] - [d_b]) \cdot [R].
\]

One would expect that the LBS upon decrypting \([z]\) obtains

\[
z = (d_a - d_b) \cdot R,
\]

and that this would only reveal the sign of \(d_a - d_b\) and not the exact value, thereby, thwarting the attack mentioned previously.

We next show that the above method is insecure too. We use the technique from [11] to recover the difference \(d_a - d_b\) from \(z\) alone with a good probability. The idea is to use the fact that every \(d_a\) and \(d_b\), \(0 \leq |d_a - d_b| \leq m\). Therefore, there will be a "small" factor of \(z\) that is less than \(m\) and, hence, feasible to recover this factor. This factor would be a possible candidate for \(R\). One could use the brute-force technique to factorize, or, for larger values of \(m\), the elliptic-curve method of factorization would be more efficient. Once a possible value of \(R\) is determined, then

\[
d_a - d_b = z/R.
\]

In case one ends up with many candidates for \(R\), then we need to brute force over these choices of \(R\), and then check the consistency of these computed differences with other similarly computed distances. This way inconsistent choices of \(R\) are eliminated. In the worst case, we may end up with more than one possibility for the distances \(d_i\) and, hence, as many possibilities for \((X_a, Y_a)\).

Hence this fix too would not lead to a secure protocol.

### 4 CONCLUSION

In this paper, we analyzed the correctness of the protocol of Ma et al. [10]. We showed that their efficient "secure" comparison technique does not give the correct output. We then showed that straightforward attempts to fix this flaw would lead to security vulnerabilities, where the location-based service provider will be able to recover information about users locations. It seems that fixing the protocol of Ma et al. is non-trivial without incurring significant cost in terms of computation time and communication bandwidth.

There have been several attempts to design a privacy-friendly comparison protocol that is significantly more efficient than the homomorphic/MPC evaluation of the entire comparison circuit. Unfortunately, many of them have been shown to be insecure. Hence, it is an interesting open problem to design a comparison protocol that is lightweight in terms of both time and bandwidth.

### ACKNOWLEDGMENTS

This work was partially funded by the Infosys Foundation Career Development Chair Professorship grant for Srinivas Vivek.

### REFERENCES

[1] Anamaria Costache, Nigel P. Smart, Srinivas Vivek, and Adrian Waller. 2016. Fixed-Point Arithmetic in SHE Schemes. In Selected Areas in Cryptography - SAC 2016 - 23rd International Conference, St. John’s, NL, Canada, August 10-12, 2016. Revised Selected Papers (Lecture Notes in Computer Science, Vol. 10532). Springer, 323–324. https://doi.org/10.1007/978-3-319-69453-5_22

[2] Ivan Damgård, Martin Geisler, and Mikkel Krøigaard. 2007. Efficient and Secure Comparison for On-Line Auctions. In Information Security and Privacy, 12th Australasian Conference, ACISP 2007, Townsville, Australia, July 2-4, 2007, Proceedings (Lecture Notes in Computer Science, Vol. 4586). Josef Pieprzyk, Hossein Ghodosi, and Ed Dawson (Eds.). Springer, 416–430. https://doi.org/10.1007/978-3-540-73458-1_30

[3] Ivan Damgård, Martin Geisler, and Mikkel Krøigaard. 2009. A correction to ‘efficient and secure comparison for on-line auctions’. Int. J. Appl. Cryptogr. 1, 4 (2009), 323–324. https://doi.org/10.1504/IJACT.2009.028031

[4] Junfeng Fan and Frederik Vercauteren. 2012. Somewhat Practical Fully Homomorphic Encryption. IACR Cryptol. ePrint Arch. 2012, 144 (2012). 144. http://eprint.iacr.org/2012/144
Revisiting a Privacy-Preserving Location-based Service Protocol using Edge Computing

1. Efficient Sorting of Homomorphic Encrypted Data With k-Way Sorting Network. IEEE Trans. Inf. Forensics Secur. 16 (2021), 4389–4404. https://doi.org/10.1109/TIFS.2021.3106167

2. Messages in a Concealed Bottle: Achieving Query Content Privacy With Accurate Location-Based Services. IEEE Transactions on Vehicular Technology 67, 8 (2018), 7698–7711. https://doi.org/10.1109/VT.2018.2838041

3. Location-based services. Commun. ACM 51, 3 (2008), 65–69.

4. Revisiting Driver Anonymity in ORide. In Selected Areas in Cryptography - 28th International Conference, SAC 2021, Virtual Event, September 29 - October 1, 2021, Revised Selected Papers (Lecture Notes in Computer Science, Vol. 13203), Riham AlTawy and Andreas Hülsing (Eds.). Springer, 2021, 25–46. https://doi.org/10.1007/978-3-030-99277-4_2

5. Moving average — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/wiki/Moving_average. [Online; accessed 08-May-2022].

6. Continuous K Nearest Neighbor Query Scheme with Privacy and Security Guarantees in Road Networks. In 2018 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trust Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). IEEE, 2018, 828–833. https://doi.org/10.1109/SmartWorld.2018.00153