A Privacy protection scheme for carpooling service using fog computing

Bangyan KOU, Suzhen CAO, Jia LV
Department of Computer Science and Engineering College of Northwest Normal University Lanzhou, China
15117291712@163.com, 576342353@qq.com, Lvjiajia_2019@163.com

Abstract—As the traffic environment becomes more and more crowded, an increasing number of people are willing to carpool. Carpooling services provide convenience for drivers and passengers to share their journeys. But carpooling requests are uploaded to cloud server for processing and analysis in most existing schemes, resulting in communication delays and leakage of user privacy. To solve this problem, we propose a privacy protection scheme for carpooling service using fog computing. First, identity-based digital signature technology is used to achieve anonymous authentication. Second, location tags, bloom filters and fuzzy extractors are utilized to match passengers and drivers without disclosing the users’ location; Consortium blockchain technology is adopted to store carpooling records. Finally, security analysis and experiments results show that our scheme offers privacy preservation and security, while maintaining low computational cost.

1. INTRODUCTION
In recent years, with the rise of the shared economy, carpooling service has gradually become a green fashion. Carpooling service can not only save passenger travel time and reduce traffic congestion, but also reduce vehicle emissions and noise. There are many online carpooling service providers like Uber and Didi around the world. According to [1], the ride-sharing market is expected to reach US$2205 billion by 2025. Carpooling services involve sensitive information such as the user’s identity and location. To protect the privacy information, users use anonymous identities and encrypt their location data. However, anonymization and encryption make it difficult for the carpooling system to match passengers and drivers’ on and off location. Sherif et al. [2] proposed a privacy-preserving Ride Sharing Scheme for Autonomous Vehicles in Big Data Era, which uses group signatures to ensure user anonymity and utilizes similarity measurement techniques to match users. Pham et al. [3] put forward a privacy-preserving yet accountable ride-hailing service, which exposes the passenger's ride location to the server. And the user must request an anonymous certificate from the server every time rides, causing complicated certificate management. He et al. [4] proposed a privacy protection scheme for partner selection in ride sharing, which combines TTS and spatial regions for matching. Luo et al. [5] presented a privacy-preserving ride matching scheme, which uses road network embedding technology and homomorphic encryption to estimate the shortest distance between the drivers and passengers to match.

The above schemes require users to upload carpooling requests to remote cloud servers for matching. It will cause problems such as communication delays, computational complexity, and increased costs. Li et al. [6] proposed an efficient and privacy-preserving carpooling using blockchain-assisted vehicular Fog Computing, which uses fog nodes to process carpooling data, and it stores carpooling records and records hash values on cloud servers and private blockchain respectively. It provides data auditability...
when disputes arise between passengers and drivers. However, the scheme that the RSU is attacked is not considered. And the cloud server is semi-trusted, if the server is crashed or the data is tampered, it cannot provide reliable data auditability.

To solve the above problem, we propose a privacy protection scheme for carpooling service using fog computing. First of all, we use pseudo-identity to achieve user anonymity and traceability, and adopt Bloom filter to represent user’s location label, current location and destination to achieve passenger and driver location matching. Then, we utilize fuzzy extractor to extract the shared secret key between matching location tags [7]. Last, Fog nodes build consortium blockchain [8] to store carpooling records, which solves the security problem of centralized storage and ensures data auditability.

2. PRELIMINARIES

2.1. Bilinear Mapping
Let $G_1$ and $G_2$ be two cyclic addition groups and cyclic multiplication groups of prime $q$, $g$ is the generator of $G_1$, $e : G_1 \times G_1 \rightarrow G_2$ is the bilinear mapping, which satisfies the following three properties:

- Bilinearity: For $a, b \in \mathbb{Z}_q^*$, $e(g^a, g^b) = e(g, g)^{ab}$.
- Non-degeneracy: $e(g, g) \neq 1$.
- Computability: For any $g_1, g_2 \in G_1$, it is an effective algorithm to calculate $e(g_1, g_2)$.

2.2. Difficult Problem
Computational Diffie–Hellman Problem (CDHP): $G_1$ is a cyclic group of order $q$, $g$ is the generator of $G_1$, it is difficult to calculate $g^{ab} \in G_1$ when $(g, g^a, g^b) \in G_1$ is known.

3. SYSTEM MODEL
The system model contains five participants: $TA$, consortium blockchain, $FN$, passengers and drivers. The system model diagram of this paper is shown in Figure 1.

- $TA$: It is responsible for the registration of the $FN$, passengers and drivers. In order to protect the privacy of passengers and drivers, $TA$ generates a pseudo identity for each user. When dispute occurs to trace the violating user, $TA$ can present or expose the real identity of the user and sanction the user.
- Consortium blockchain: It is a decentralized, safe and reliable database maintained by all fog nodes. It is responsible for storing the carpooling records and ensuring the auditability of the data.
- $FN$: This is a fog node with computing and communication functions, which is deployed on the edge of networks. It is responsible for collecting carpooling requests and responses in the covered area.
to matching passengers and drivers. After a successful match, it is in charge of storing the carpooling records in the consortium blockchain.

- **Passenger:** It is responsible for sending a carpooling request to the local \( F_N \), including the pseudo-identity, start time, end time, ciphertext of the current location and destination, number of passengers, and reputation value.
- **Driver:** It is in charge of sending a carpooling response to the local \( F_N \), including the pseudo identity, ciphertext of the current location and destination, the number of vehicle vacancies, and reputation value.

4. PROPOSED SCHEME

(1) System Initialization

\( TA \) selects a security parameter \( k \), two cyclic groups \( G_1 \) and \( G_2 \) with prime \( q \), a generator \( g \) of \( G_1 \) and a bilinear map \( e : G_1 \times G_1 \rightarrow G_2 \), a master key \( s \in \mathbb{Z}_q^* \), three hash functions:

\[ \begin{align*}
H_1 : G_1 \times \{0,1\} &\rightarrow \mathbb{Z}_q^*, \\
H_2 : \{0,1\} &\rightarrow \mathbb{Z}_q^*, \quad H_3 : \{0,1\} \rightarrow G_1.
\end{align*} \]

Then, \( TA \) random chooses a filtering function of length \( n \), and computes \( P_{pub} = g^s \). Finally, \( TA \) publishes the system parameters \( \{k, P_{pub}, G_1, G_2, g, q, H_1, H_2, H_3, n, \phi\} \).

(2) Registration

① When a passenger with identity \( iU \) participates the carpooling system, \( TA \) calculates \( 1(, ) \) \( iU i pub P ID U H P s \triangleq \), \( 2(, ) \) \( iU iUSH H P I D s s + \) as pseudo-identity and private key of \( iU \), and uses the master key to sign initial reputation value \( i_r \) of the passenger.

② When a driver with identity \( jD \) participates carpooling system, \( TA \) calculates \( 1(, ) \) \( jD j pub P ID D H P s \triangleq \), \( 2(, ) \) \( jD jDDSH H P I D s s + \) as pseudo identity and private key of \( jD \), and uses the master key to sign initial reputation value \( j_r \) of the driver.

③ Each fog node \( xF_N \) needs to register with \( TA \). \( TA \) computes \( 2(, ) \) \( xF N xSH H P I D s s + \) as private key of \( xF_N \).

(3) Carpooling Request

① \( iU \) selects a set of environment signals \( 12(, ) \) \( iX tt \), \( t_1 \) and \( t_2 \) are the start time and end time. \( iU \) extracts the location tag \( 12 12(, ) \) \( iY tt X tt \triangleq \phi \), and inserts it into bloom filter \( iw \) by applying \( 21 2((, ) , ) \) \( ii iw H Y t t w \triangleq \).

② \( iU \) calculates a temporary key \( 12(, ) \) \( i s kp krr \), and embeds the public key \( ip kr \) in the location tag using the code offset construction [9] of fuzzy extractors proposed by Dodis, et al, \( iip kc Encode n l r \), \( iii s hcw \triangleq \).

③ \( iU \) converts the current position \( iloc \) and destination \( igo \) into the minimum grid block set \( 12(, ) \) \( ig loc ig go \), \( iU \) embeds them into bloom filters \( \hat{\nu}_1 \) and \( \hat{\nu}_2 \) by applying \( \hat{\nu}_i = \text{Inst}(H_2(g(loc), \nu_i), \nu_i) \), \( \hat{\nu}_i = \text{Inst}(H_2(g(\nu_i), \nu_i), \nu_i) \), and sets the number of passengers \( EM_i \).

④ \( iU \) forms carpooling requests \( 12(, ) \) \( RE = \{ID_i, t_1, t_2, \hat{\nu}_1, \hat{\nu}_2, sh_i, EM_i \} \), and computes signature \( \sigma_i = (v_i, Q_i) \), \( Q_i = g^{v_i}, h_i = H_3(RE, Q_i, PID_i) \), \( v_i = h_i^{s_{\nu_i}} \), as well as sends \( \{ID_i, RE_i, \sigma_i\} \) to the local fog node \( FN_i \).

(4) Fog Node Response

After receiving the passenger’s carpooling request, local fog node \( FN_i \) verifies whether equation \( \hat{e}(v_i, g) = \hat{e}(Q_i, h_i) \) is true. And if it is true, \( FN_i \) broadcasts message \( S_i = \{t_1, t_2, sh_i\} \) in its coverage area in the order of reputation value \( r_i \) from high to low.
(5) Carpooling Response

1. After receiving the message $S_{ij}$, $D_j$ calculates location tag $Y_j(t_i,t_2) = \varphi(X_j(t_i,t_2))$, and inserts it into the Bloom filter $w_j$.

2. $D_j$ extracts public key $r_{pk}$ by applying $c_j = sh_i + w_j$, $r_{pk} = Decode(n,i,c_j)$.

3. $D_j$ covert the current position $loc_j$ and destination $go_j$ into the minimum grid block set $gl_j$ and $gg_j$, as well as encrypts them with $ipkr$ to get $H_1(gl_j \| r_{pk})$ and $H_2(go_j \| r_{pk})$. $D_j$ sets the number of passengers $Em_j$.

4. $D_j$ forms carpooling response $RE_j = \{PID_j, EM_j, H_1(gl_j \| r_{pk}), H_2(go_j \| r_{pk})\}$ and computes signature $\sigma_j = (v_j, Q_{h_j}), Q_{h_j} = g^{s_{v_j}}, h_j = H_1(RE_j, Q_{h_j}, PID_j), v_j = h_j^{s_{v_j}}$, as well as sends $\{PID_j, RE_j, \sigma_j\}$ to the local fog node $FN_i$.

(6) Carpooling Matching

After receiving carpooling response, local fog node $FN_i$ verifies whether equation $\hat{\epsilon}(v_j, g) = \hat{\epsilon}(Q_{h_j}, h_j)$ is true. If it is true, $FN_i$ uses the bloom filter query function to match the current location and destination in the order of reputation value $r_j$ from high to low, $Que(H_1(gl_j \| r_{pk}), \bar{w}_{ij})$, $Que(H_2(go_j \| r_{pk}), \bar{w}_{ij})$. After matching, $FN_i$ judges whether $EM_i < EM_j$ is true. If it is true, $FN_i$ will send $PID_j$ and $PID_{ij}$ to passenger and driver respectively. Otherwise, $FN_i$ rejects the driver's carpooling response. After receiving the message, the driver and the passenger use $\{r_{pk}, r_{pk}\}$ to communicate the specific carpooling information.

$FN_i$ signs the carpooling record $CO_{ij} = \{RE_i, RE_j\}$ and packs it into data blocks. Block headers contains the signature of $FN_i$, the hash value of the current block, the hash value of the previous block, time stamp, Merkle root. $FN_i$ uses PBFT mechanism [10] to broadcast transactions to all fog nodes in the whole system to obtain consensus. The steps of consensus are as follows:

1. Pre-preparation stage: Primary node $FN_i$ broadcasts transactions to other fog nodes. The total number of fog nodes is $t$, and the number of abnormal nodes is $f$.

2. Preparation stage: After non-primary nodes $FN_i$ receives the transaction, it reviews the integrity and legitimacy of the transaction, and signs and broadcasts the audit result to other non-primary nodes. The preparation phase has been completed when the node receives messages from $2f$ different nodes.

3. Commit stage: The node summarizes the audit results of other non-primary nodes and compare them with its own audit results. If the node receives $t-f$ (including its own messages) confirmation messages, it sends the results to the primary node.

4. When all fog nodes reach consensus, the master node $FN_i$ adds new blocks to the consortium blockchain. $FN_i$ analyzes the unconfirmed nodes in the audit results, determines whether these nodes are malicious, and handles abnormal nodes in time. $TA$ will reject abnormal nodes according to the feedback results.

(7) Carpooling Termination and Tracking

If the user exits from the carpooling matching process before the matching is successful, the user sends canceling carpooling message to $FN_i$. After receiving the message, $FN_i$ will delete the carpooling information of the user from the waiting carpooling list. If the user exits from the carpooling matching process after the matching is successful, the passenger needs to send cancellation message to $FN_i$ and the driver who successfully matches.

If the user has cheated or failed verification during the carpooling process, the true identity of the user can be exposed after investigation and confirmation of user's cheating behavior.
5. SECURITY ANALYSIS AND PROOF

5.1. Security Analysis

(1) Identity Authentication and Data Integrity
In the scheme, FN, performs identity authentication and data integrity verification to ensure that the data are sent by the legitimate user and has not been modified. It can be seen from Theorem 1 that since the CDH problem is difficult, no adversary can generate legal carpooling requests and carpooling responses. Therefore, FN can perform user identity authentication and check the integrity of carpooling requests and carpooling responses by verifying whether equation \( \hat{\delta}(v, g) = \hat{\delta}(Q, h) \) is true.

(2) Anonymity
When users join the carpooling system, TA generates pseudo-identity \( U_i \) for the user. During carpooling process, there will be no third party other than TA who knows the user's true identity. The anonymity requirement is satisfied, and the identity privacy of users is guaranteed.

(3) Data Confidentiality
① \( U_i \) embeds location tags and public key \( r_{pk} \) in the bloom filter \( w_i \). Since FN is not near \( U_i \), it cannot obtain public key \( r_{pk} \) of the \( U_i \), and use \( r_{pk} \) to query \( \hat{w}_1 \) and \( \hat{w}_2 \).
② According to the nature of BCH decoder, no \( D_j \) can judge whether he has extracted the correct \( r_{pk} \) or he is in the location tags of \( U_i \).
③ Current location and destination of the \( D_j \) are calculated as \( H_j(g(loc)) \| r_{pk} \) and \( H_j(g(go)) \| r_{pk} \), so FN cannot obtain location information of the \( D_j \) without knowing \( r_{pk} \).
④ FN does not send \( H_j(g(loc)) \| r_{pk} \) and \( H_j(g(go)) \| r_{pk} \) to \( U_i \), so \( U_i \) could not get the driver's position information.

(4) Location Authentication
If \( loc_i \) and \( loc_j \) are different, and \( D_j \) successfully extracts public key \( r_{pk} \) of the \( U_i \), \( D_j \) is considered to launch location cheating attack [9]. Although the difference between \( loc_i \) and \( loc_j \) is greater than the threshold value \( T \), due to the false positives of the bloom filter, the formula \( Ham(c, c') \leq er \) may be established (representing the maximum number of bits that can be corrected by BCH coding), which leads to the successful extraction \( r_{pk} \) by \( D_j \) far from \( U_i \), but \( U_i \) can reduce this situation by increasing the size of the bloom filter.

(5) Traceability
If the carpooling user has violations or signature verification failures during the carpooling process, TA uses the master key to calculate \( U_i = PID_{U_i} \oplus H_i(s, P_{pub}) \), and exposes the real identity of the user. The scheme does not need to store any anonymous certificates, which saves the storage cost.

(6) Decentralization
Using the distributed storage method based on consortium blockchain, not only does not rely on the trusted third-party database, but also reduces the cost of maintaining the centralized database, and avoids the threat of the traditional centralized data storage vulnerable to malicious attacks.

5.2. Security Proof
If the CDH problem is difficult, the proposed scheme in this paper can resist the existence forgery attack under the adaptive selection message attack under the random prediction model. It can be proved by the following theorem.

Theorem 1. In the random oracle model, if there is an attacker \( A \) who can win in time \( t \) by \( e \) after \( q_{Hi} \) time \( H_i \) queries, \( q_{hi} \) time \( H_i \) queries, \( q_{Kg} \) time \( Key - Gen \) queries, \( q_s \) time \( Sign \) queries. The time of one
query is respectively $t_1, t_2, t_K, t_S$. Then a challenger $C$ can solve the difficult problem of CDH with the advantage of $\varepsilon' \geq (1 - \frac{1}{q_{t_1}})(1 - \frac{1}{q_{t_2}})(1 - \frac{1}{q_{t_K}})(1 - \frac{1}{q_{t_S}})(1 - \frac{t'}{2^t})$ within the time $t' < t + q_{t_K}t_{Kg} + q_{t_S}t_S + 2q_{t_1}t_{1h} + 2q_{t_2}t_{2h}$.

**Proof.** $C$ gives an example $(g, g^a, g^b)$ of the CDH problem example, and generates system parameters. return $\{ k, P_{pub}, G_1, G_2, g, q, H_1, H_2 \}$ to $A$. $C$ saves the master key $s$, selects challenge pseudo identity $PID_{U_i}$, and returns $\{ k, P_{pub}, G_1, G_2, g, q, H_1, H_2 \}$ to $A$. $A$ initiates the following queries: $H_i$ query: $A$ performs a query on $PID_{U_i}$ if $L_{H_i}$ contains $(PID_{U_i}, c_i)$, then $C$ returns $c_i$ to $A$. Otherwise, $C$ performs the following operations: (1) If $PID_{U_i} = PID_{U_i'}$, $C$ aborts the query. (2) If $PID_{U_i} \neq PID_{U_i'}$, $C$ selects $c_i \in Z_q$, returns $c_i$ to $A$, and adds $(PID_{U_i}, c_i)$ to list $L_{H_i}$.

Key-Gen query: $A$ performs a query on $PID_{U_i}$. If $L_{key}$ contains $(PID_{U_i}, S_{U_i})$, then $C$ returns $S_{U_i}$. Otherwise, $C$ performs the following operations: (1) If $PID_{U_i} = PID_{U_i'}$, $C$ aborts the query. (2) If $PID_{U_i} \neq PID_{U_i'}$, $C$ queries $L_{H_i}$ get $(PID_{U_i}, c_i)$, computes $S_{U_i} = c_i + s$, then returns $S_{U_i}$ to $A$ and adds $\{PID_{U_i}, S_{U_i}\}$ to list $L_{key}$.

$H_2$ query: $A$ performs a query on $PID_{U_i}$. If $L_{H_2}$ contains $(PID_{U_i}, m_i, Q_{iU}, h_i)$, then $C$ returns $h_i$ to $A$. Otherwise, $C$ performs the following operations: (1) If $PID_{U_i} = PID_{U_i'}$, $C$ computes $Q'_{iU} = g^{h_i}$, $h'_i = g^a$, returns $h'_i$ to $A$ and adds $(PID_{U_i}, m_{iU}, g^{h_i}, g^s)$ to list $L_{H_i}$. (2) If $PID_{U_i} \neq PID_{U_i'}$, $C$ queries $L_{h_i}$ get $(PID_{U_i}, S_{U_i})$, computes $Q_{iU} = g^{S_{U_i}}$. Then, $C$ randomly selects $r_i \in Z_q$, computes $h_i = g^{r_i}$. $C$ returns $h_i$ to $A$ and adds $(PID_{U_i}, m_{iU}, g^{h_i}, g^{r_i})$ to list $L_{H_i}$.

Sign query: $A$ performs a query on $\{PID_{U_i}, m_{iU}\}$. $C$ queries $L_{H_i}$ get $h_i$, performs the following operations: (1) If $PID_{U_i} = PID_{U_i'}$, $C$ aborts the query. (2) If $PID_{U_i} \neq PID_{U_i'}$, $C$ queries $L_{key}$ get $(PID_{U_i}, S_{U_i})$, computes $v_i = h_i^{r_i}$, and returns $v_i$ to $A$.

Finally, adversary $A$ stops query and outputs the valid signature $v_i$ of user $PID_{U_i}$ to $m_{iU}$. If $PID_{U_i} \neq PID_{U_i'}$, $C$ outputs false. If $PID_{U_i} = PID_{U_i'}$, $C$ queries $L_{H_i}$ get $Q'_{iU} = g^{h'_i}$, $h'_i = g^a$. Since the signature is legal, that is $\hat{e}(v_i, g) = \hat{e}(Q'_{iU}, h'_i) = \hat{e}(g^a, g^b)$, then $g^{ab} = v_i$. Challenger $C$ solves the difficult problem of CDH.

6. **PERFORMANCE ANALYSIS**

The experimental environment is Inter Core i7, 2.5GHz, Windows 10 operating system. Based on the PBC database of 0.4.7, the scheme of this article, [6] and [11] were tested. The computational costs of 10-100 users in the registration stage, carpooling request stage, carpooling response stage and carpooling matching stage are simulated.

Figure 2 shows the calculation cost of our proposed scheme, [6] and [11] in the registration stage. Figure 3 represents the calculation cost of our proposed scheme, [6] and [11] in the carpooling request stage. Figure 4 demonstrates the calculation cost of our proposed scheme, [6] and [11] in the carpooling response stage. Figure 5 illustrates the calculation cost of our proposed scheme, [6] and [11] in the carpooling matching stage. With the increase of the number of users, the time required for each stage is also gradually increases. However, compared with the schemes in [6] and [11], the computational cost of our proposed scheme is the least in each stage.
7. CONCLUSION
Aiming at the security and privacy problems in carpooling service, this paper proposes a privacy protection scheme for carpooling service using fog computing. Anonymous authentication and data integrity can be ensured by digital signature technology. The proposed scheme uses location tag, fuzzy extractor and bloom filter to perform carpooling matching, providing low waiting time service without disclosing the privacy of users. Consortium blockchain is built to store carpooling records to solve the security risk of centralized data storage. Security analysis shows that the scheme meets the requirements of security and privacy. The simulation results demonstrate that the proposed scheme is more effective. The next step is to achieve a navigation service privacy protection scheme for fog computing.

ACKNOWLEDGMENT
This work was supported by the National Natural Science Foundation of China (61662069, 61662071, 61562077).
REFERENCES

[1] Ride sharing market cap. [Online]. Available: http://www.ln632.com/132.html.2019/07/10.

[2] A Sherif, K Rabieh, M Mahmoud, et al, “Privacy-Preserving Ride Sharing Scheme for Autonomous Vehicles in Big Data Era,” IEEE Internet of Things Journal, 2017, vol. 4, no. 2, pp. 611-618.

[3] A Pham, P Dacosta, I Italo, et al, “ORide: A privacy-preserving yet accountable ride-hailing service,” Proc 26th USENIX Security Symp, 2017, pp. 1235-1252.

[4] Y He, J Ni, X Wang, et al, “Privacy-preserving Partner Selection for Ride-sharing Services,” IEEE Transactions on Vehicular Technology, 2018, pp. 1-1.

[5] Y Luo, X Jia, S Fu, M Xu, “pRide: Privacy-Preserving Ride Matching Over Road Networks for Online Ride-Hailing Service,” IEEE Transactions on Information Forensics and Security, 2019, vol. 14, no. 2, pp. 1791-1802.

[6] M Li, L Zhu, X Lin, “Efficient and Privacy-Preserving Carpooling Using Blockchain-Assisted Vehicular Fog Computing,” IEEE Internet of Things Journal, 2019, vol. 6, no. 3, pp. 4573-4584.

[7] Y Zheng, M Li, W Lou and Y T Hou, “Location Based Handshake and Private Proximity Test with Location Tags,” IEEE Transactions on Dependable and Secure Computing, 2017, vol. 14, no. 4, pp. 406-419.

[8] D Wang, X Zhang, “Secure Data Sharing and Customized Services for Intelligent Transportation Based on a Consortium Blockchain,” IEEE Access, 2020, vol. 8, pp. 56045-56059.

[9] Y Dodis, R Ostrovsky, L Reyzin, A Smith, “Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data,” Proc Adv Cryptol, 2004, pp. 523-540.

[10] X Zhang, X Chen, “Data Security Sharing and Storage Based on a Consortium Blockchain in a Vehicular Adhoc Network,” IEEE Access, 2019, pp. 1-1.

[11] J Ni, K Zhang, X Lin, H Yang, X Shen, “AMA: Anonymous mutual authentication with traceability in carpooling systems,” IEEE Int Conf Commu, 2016, pp. 1–6.