IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 8, 2020, accepted April 24, 2020, date of publication April 28, 2020, date of current version May 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2990932

A Top-*K* Query Scheme With Privacy Preservation for Intelligent Vehicle Network in Mobile IoT

CHANGLI ZHOU[1,2], **(Member, IEEE), TIAN WANG**[1,2], **(Member, IEEE), HUI TIAN**[1], **WENXIAN JIANG**[1], **AND ZHIJIAN WANG**[3]

[1]College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China
[2]Key Laboratory of Data Mining and Intelligent Recommendation, Fujian Province University, Xiamen 361021, China
[3]Information Science School, Guangdong University of Finance and Economics, Guangzhou 510320, China

Corresponding author: Changli Zhou (zcl@hqu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61802134, and in part by the Open Fund of Key Laboratory of Data Mining and Intelligent Recommendation, Fujian Province University under Grant DM201905.

ABSTRACT The emergence of mobile Internet of Things (IoT) has brought much more convenience to our daily life than ever before, modern intelligent vehicles are usually equipped with various mobile IoT components or devices. However, when we are enjoying the convenience from the rapid development of mobile IoT services, our privacy can be misused by attackers in an easier way. In this paper, we devise a query scheme with intelligent vehicle privacy guarantees, which enables a vehicle to acquire accurate query service from the service provider (SP) without providing its explicit private information, such as location or query interest privacy. And more than that, we introduce network coding for the first time to make our scheme applicable to a more sophisticated top-*K* query by multiple vehicles cooperation in mobile IoT. Furthermore, we also consider the unnecessary reveal issue of service data at SP side since the data is its asset. Performance analysis and experiments verify the validity of our scheme and demonstrate a better accuracy and efficiency compared with existing solutions in mobile IoT scenario.

INDEX TERMS Mobile Internet of Things (IoT), intelligent vehicle network, privacy preservation, top-K query, oblivious transfer (OT), network coding (NC), private information retrieval (PIR).

I. INTRODUCTION

With explosively emerging of mobile Internet of Things, numerous smart components are equipped on mobile IoT terminals, such as intelligent vehicles, wearable electronics and smart phones, these mobile terminals usually cooperate with each other to collect, process and transmit data of users, which is the most popular application in our daily life [1].

There are three types of entities in mobile IoT framework: intelligent terminals as we mentioned above, intermediate nodes which carry out data aggregation and preliminary processing, and service providers (SP) which respond the service requests from intelligent terminals, as shown in Fig.1.

Intelligent vehicle network which utilizes mobile terminals are becoming one of the emerging applications in mobile IoT, and the mobile services in intelligent vehicle networks have given rise to widespread interests due to its convenience and effectiveness [1], [2], such as location-based service, navigation, querying for the *K* nearest neighbor (*K*NN) places of interest (PoI), etc. With a variety of sensors or smart components inside a vehicle, diverse and plenty of personal data can be collected and utilized, such as location, velocity, voice, image and even fingerprint. More than that, nearly all of the mobile services ask intelligent vehicles to provide sensing or private data for generating accurate service results, which may extremely lead to privacy exposure of these vehicles.

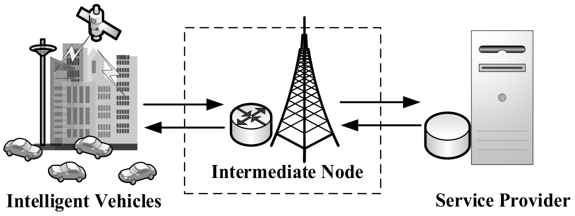


FIGURE 1. The entities in mobile IoT framework.

The associate editor coordinating the review of this manuscript and approving it for publication was Weizhi Meng.

Malicious attackers are capable to infer deeper private information from a mobile service request associating with background knowledge [2]–[5], such as the financial situation, hobbies, health status, and home address. Therefore, when we enjoy the convenience from intelligent vehicles, we must take effective measures to preserve the sensing and private data collected by the smart IoT components inside a vehicle. For sake of simplicity, we pick location and query interest as two typical privacy of an intelligent vehicle, and devise a service query scheme with privacy preservation for the intelligent vehicle network in mobile IoT scenario.

A. LOCATION PRIVACY PROTECTIONS IN MOBILE IoT

Several remarkable overviews have summarized the research focuses of location privacy in mobile IoT and potential hotspots in the future [2], [6]–[8]. Normally, there are mainly two categories of location protections in location-based mobile IoT: cloaking region [3], [4], [9], [10] and dummy [11]–[14].

On the one hand, cloaking region refers to an area which is employed by different mobile terminals to blur their sensing location in a query, it is generally constructed by a vehicle or a trusted third-party server. However, owing to the route uncertainty of all involved vehicles, it's hard to construct successive cloaking regions maintaining k-anonymity in mobile IoT [9], i.e., successive cloaking regions should retain most vehicles in the initial cloaking region, such that attackers are uncertain to find the actual vehicle who initiates the query. And most trusted servers confront a service bottleneck at the demands of cloaking region construction [10], [15]. Wang et al have discussed the balance between privacy preservation and data integrity in various IoT and the trust issues of data collection [16]–[23]. Arain et al have proposed a remarkable location protection approach and a dynamic pseudonymous authentication protocol based on mix-zones, which are practical and efficient in real road network scenario [24], [25].

On the other hand, dummies are pseudo locations instead of the actual sensing location by sensors in a mobile service request, dummies are usually generated around the actual location by smart components [11]. A typical service request with dummy is first proposed by Yiu [12], which is a valid solution to guarantee location privacy, but it has the issues of inaccuracy, failing to achieve k-anonymity and location diversity. To deal with these issues, Zhou has improved the service accuracy with dummy and presented a query scheme in mobile IoT scenario [11]. Gong and Huang [13], [14] have made significant improvements of Yiu's work in the aspect of achieving k-anonymity. But most of these protections are devised in the Euclidean space without considering building blockage, which leads to inaccurate query result, little work has paid attention to the real scenario in mobile IoT.

B. INTEREST PROTECTIONS IN MOBILE IoT

Different from those sensitive sensing data captured by smart components, such as vehicle location, speed, query interest directly reflects vehicle privacy, while most service providers ask vehicles to provide their interests for high-quality service. Trustworthy services should always have as little access to private data as possible without sacrificing service quality.

An existing effective protection for user interest privacy is private information retrieval (PIR) [3], [4], [15], [26]–[28]. By means of PIR, vehicles are capable to calculate desired query results after obtaining a response cryptographic data block from SP without revealing any private information. However, most existing PIR schemes are devised on strong cryptology and inapplicable to mobile IoT due to expensive processing overhead from frequent service requests with constantly updating sensing data [15], [26]. Moreover, most PIR schemes have to utilize a trusted third party [27], [28] between a vehicle and an SP to preprocess the database, which increases the risk of data exposure.

Remarkably, two practical PIR schemes [3], [4] have been devised in Euclidean space with divided cells, both of which make good use of the homomorphic properties [29], [30] of public key cryptosystems [28], [31], but they obviously have the distance inaccuracy issue due to building blockage in a real mobile scenario, and the vehicle can be identified in a particular cell risking location privacy disclosure, and querying in a cell also gives rise to more extensive searching area with too much additional processing overhead.

Besides, intelligent vehicles always have to repeatedly send service requests to the SP as long as it enters into a new cell for obtaining the latest service results, so the vehicle may get a great deal of reduplicative service results which have been obtained in the former cell already, that's a waste of bandwidth. We will consider these mentioned issues and devise a novel and practical PIR query scheme for intelligent vehicle network in mobile IoT.

C. CHALLENGES OF TOP-K PIR QUERY IN MOBILE IoT

Specifically, we take the most common PoI query from intelligent vehicles as an example to illustrate our query scheme. Currently, a majority of query schemes in mobile IoT focus on the KNN PoI query service [5], [27], while most vehicles always expect to retrieve the most popular PoIs which are ranked by the click rate, comments and etc., i.e., top-K query [26], [32], [33], rather than only ranked by distance in the KNN scheme. Therefore, the SP should record all the query interests from the intelligent vehicles, and periodically rank the PoI records according to the popularity.

However, it means the SP needs to be aware of the vehicle interests in a top-K query, but it is opposite to the core idea of PIR which facilitates a vehicle to receive desired service without disclosing its privacy. It's absolutely a great challenge to achieve top-K query with PIR which provides no vehicle interest [27]. To the best of our knowledge, this is the first paper to deal with this issue. We will achieve top-K query maintaining PIR by introducing network coding (NC) [34] with intelligent vehicle cooperation in mobile IoT scenario based on our previous work [11].

NC has been formerly considered as a promising data dissemination approach to improve the network

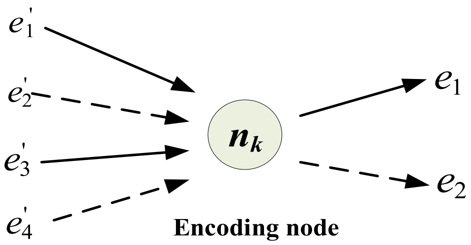


FIGURE 2. An encoding node in networks.

performance [35], [37]. A forwarding node in networks no longer follows traditional storage and forward mode, it combines the input packets into one output packet, and the receiver could calculate the original data with these combined packets in Fig.2. There is a compelling characteristic of NC, i.e., the receiver is capable to calculate the original data iff enough encoded packets are collected, we will take advantage of this beneficial characteristic to prevent SP from acquiring vehicle interest confidently for privacy concerns. There are also security and privacy solutions with NC [36], [37], but they are not applicable to deal with these issues. It's the first time to introduce NC to achieve the top-K query in mobile IoT.

To sum up, for private data preservation which is collected and utilized by intelligent vehicles in mobile IoT, we summarize the issues to be resolved as follows:

- Initially, private data collected by intelligent vehicles should be preserved well in a mobile service request without revealing privacy to any service provider (SP);
- However, SP has to be aware of the private data, such as vehicle's service interest, to calculate the popularity and then ranks the top-K service records in its database, it's hard to achieve a top-K query with privacy preservation;
- Most solutions for privacy preservation in mobile IoT are either built on strong cryptography accompanying with excessive overhead, or built in Euclidean space without considering building blockage, both of which lead to inaccurate result and impractical implementation. And SP also desires strong control on its data security.

D. CONTRIBUTIONS

To deal with the mentioned issues, we achieve a top-K query scheme with mobile IoT devices cooperation in this paper, it protects vehicle privacy in mobile IoT scenario. Our scheme is built on a strong cryptosystem for security concerns, but with less computational and communicational overhead, we summarize our contributions as follows:

- Firstly, we devise a query scheme with vehicle privacy guarantees based on oblivious transfer (OT) and private information retrieval (PIR) in vehicle network. Taking full consideration of the influential factors in the road network scenario, our scheme has lower overhead on vehicle terminals in spite of being built on strong cryptology, and its query results are more accurate than other schemes which are proposed in Euclidean space;
- And then we further achieve a more complicated top-K query based on our aforementioned query scheme by introducing NC for the first time as far as we know. In our top-K query scheme, neighbor vehicles cooperate with each other to transmit their encoded query interests, and SP couldn't decode the interest privacy for ranking popularity until enough data block are collected and couldn't relate any interest to a specific vehicle;
- We also consider the data preservation at SP side since it's the asset of SP, i.e., SP will release no extra data to the vehicles in our scheme. Besides, we did extensive performance analysis and experiments to verify the validity of our scheme in mobile IoT scenario.

II. PRELIMINARIES

A. NOTATIONS AND SYSTEM MODEL

We utilize the ElGamal algorithm [38] in our scheme and describe the encrypting operation of a message m with a public key $pk = g^x$ as $E(m, pk) = (\varepsilon, \delta) = (g^r, g^m(pk)^r)$, where the generator is g of group G, the integers x and r are randomly picked. Assume a cyclic group G_0 which is a multiplicative subgroup on a finite field F_p with an order q and a generator g, respectively. Note that p and q are both large primes which satisfies q divides $p - 1$. We also assume the prime p and q are long enough for the security reasons. It has been proved that the ElGamal algorithm satisfies the sematic security [3], [4], [38]. TABLE 1 describes the main notations used in our scheme.

TABLE 1. The main notations in our scheme.

Symbols	Description
g	the generator of a group
pk	a public key
G	a group
$\boldsymbol{X} = (x_1, x_2, \ldots, x_h)$	a data block with h packets
$y(e)$	a linear combination of packets
$\boldsymbol{C}(e)$	an encoding coefficient on current node
$\boldsymbol{g}(e)$	a global encoding coefficient
G_m	an invertible matrix of $\boldsymbol{g}(e)$
u_k	an intelligent vehicle user
v_i	a vertex in the road network
$\overrightarrow{v_m v_n}$	a directed road segment
$P_{u_k}^K$	a top-K PoI set from u_k in road distance
k_{ij}	a symmetric key to encrypt PoI details
$keyij$	a key to encrypt k_{ij}
C_{vi}	a data block starting from v_i in TABLE II
$E(\cdot)$	the ElGamal encryption
$H(\cdot)$	a hash function
$\oplus$	an exclusive or operation
$\Vert$	a concatenating operation
$\lvert < \cdot > \rvert$	the order of an element in a group

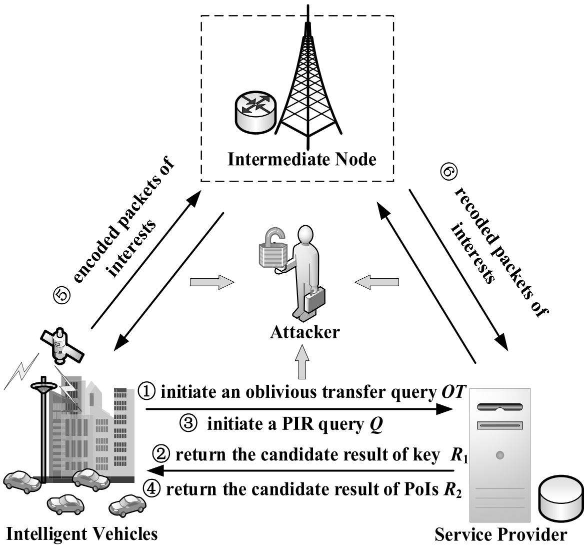

Intermediate Node
⑤ encoded packets of interests
⑥ recoded packets of interests
Attacker
① initiate an oblivious transfer query *OT*
③ initiate a PIR query *Q*
② return the candidate result of key R_1
④ return the candidate result of PoIs R_2
Intelligent Vehicles
Service Provider

FIGURE 3. Framework and query phases.

As mentioned above, there are mainly three types of entities in the system model of our scheme: intelligent vehicles, intermediate nodes and service providers, as shown in Fig.3, we describe the functionality and credibility of these entities in detail as follows:

- An intelligent vehicle collects its location and interest with the help of inside smart components, and take these private data as input to acquire mobile service from SP, but the vehicle has to preprocess its private data before transmitting it to other entities since the vehicle doesn't trust any other entity for privacy exposure concerns. The vehicle travels along the road segments and observe the traffic regulations, it has a certain of computing resource, but tries to minimize overhead to acquire the mobile service as soon as possible;
- An intermediate node is a sink node in a mobile sensor network, such as a roadside unit or a base station, it recodes the interest packets from vehicles and outputs one combination for sake of achieving the top-K query and privacy preservation in our scheme and then it transmits the combination to SP to rank the PoI records in the database of SP. The intermediate nodes are not trusted by vehicles or SP in our scheme;
- An SP responds to the query requests from vehicles, it aperiodically receives interest combinations from intermediate nodes and decodes them after getting enough ones to calculate the popularity and periodically rearrange the top-K results in its database. A distributed SP only possesses various PoI data and responds to the service requests from vehicles in its governing area. SP is not trusted by vehicles or intermediate nodes in our query scheme.

We also make the threat assumptions for attack model in our query scheme:

- Semi-honest attackers: as described in aforementioned entity credibility, there exists no trusted entity in a query except the user himself, any other vehicle, intermediate node or SP could be a semi-honest attacker, in other words, they honestly carry out the scheme while they are interested in digging deep into the identities and query interests of vehicle users for commercial interests;
- Malicious attackers: they are passive attackers, and try to infer a vehicle's location privacy and query interest when they eavesdrop the communications between the vehicle and SP, they also try to decode the interest combinations when they eavesdrop the communications between vehicles and SP through intermediate nodes. Besides, they could collude with some vehicles and launch traffic analysis attacks in the NC phase.

B. NETWORK CODING

Different from traditional forwarding mechanisms, network coding enables a set of input packets to be combined (encoded) on intermediate nodes and output one combination instead of forwarding each input, as shown in Fig.2. Assume each data block of a message transmitted from an original mobile node to a receiver can be divided into h packets $\boldsymbol{X} = (x_1, x_2, \ldots, x_h)$, an $y(e)$ indicates the linear combination of the these input packets, recorded as $y(e) = \sum_{e'} C_{e'}(e)y(e')$, wherein $\boldsymbol{C}(e) = [C_{e'}(e)]$ is the encoding coefficient on current node.

In this way, each packet in networks can be calculated by the equation $y(e) = \sum_{i=1}^{h} g_i(e)x_i$ to be a linear combination of the original packets $\boldsymbol{X} = (x_1, x_2, \ldots, x_h)$ from the original generating node, the vector $\boldsymbol{g}(e) = [g_1(e), g_2(e), \ldots, g_h(e)]$ represents a global encoding coefficient which is calculated successively by each intermediate node with distinct local encoding coefficients by the equation $\boldsymbol{g}(e) = \sum_{e'} C_{e'}(e)g_i(e')$. Suppose a receiver acquires a combination set from different routes $(y(e_1), y(e_2), \ldots, y(e_h))$, it can be expanded with an operational matrix of original packets $\boldsymbol{X}$ in Equation (1):

$$\begin{bmatrix} y(e_1) \\ \vdots \\ y(e_h) \end{bmatrix} = \begin{bmatrix} g_1(e_1) & \cdots & g_h(e_1) \\ \vdots & \ddots & \vdots \\ g_1(e_h) & \cdots & g_h(e_h) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_h \end{bmatrix} = \mathbf{G}_m \begin{bmatrix} x_1 \\ \vdots \\ x_h \end{bmatrix} \tag{1}$$

Note that, once enough combinations are received, the receiver can decode original packet $\boldsymbol{X}$ iff G_m is invertible, i.e., the matrix G_m is linear independence. It has been proved that if the encoding coefficients are chosen in a finite field $F_q(2^{16})$ and all the intermediate nodes perform random linear coding on their inputting packets to output one combination, the matrix G_m, can be full rank with a probability close to 1 [36], which is crucial to decode at SP.

The invertibility of G_m are also discussed in [37]. This characteristic of network coding can be applied to privacy preservation, i.e., only enough combinations of original vehicle interest packets are received, the SP can further decode what types of interests have been queried by vehicles without inferring vehicle privacy as a result of the decoding time delay and the resist to traffic analysis. But NC cannot be applied to

TABLE 2. A beneficial data structure to organize PoIs.

Vertex	Neighbor vertexes	PoI Type	Unique code	Top-K PoI coordinates	Top-K PoI details
v_8	$\{v_7, v_5, v_9, v_{13}\}$	Super market	1	$\{\text{cod}1,\text{cod}2,....,\text{cod}_{Kmax}\}$	$\{\text{det}1,\text{det}2,....,\text{det}_{Kmax}\}$
		Clinic	2	$\{\text{cod}1,\text{cod}2,....,\text{cod}_{Kmax}\}$	$\{\text{det}1,\text{det}2,....,\text{det}_{Kmax}\}$
		Park	3	$\{\text{cod}1,\text{cod}2,....,\text{cod}_{Kmax}\}$	$\{\text{det}1,\text{det}2,....,\text{det}_{Kmax}\}$
		......	......	......	......
v_5	$\{v_4, v_2, v_{10}, v_8\}$	Subway station	1	$\{\text{cod}1,\text{cod}2,....,\text{cod}_{Kmax}\}$	$\{\text{det}1,\text{det}2,....,\text{det}_{Kmax}\}$
		Saloon	2	$\{\text{cod}1,\text{cod}2,....,\text{cod}_{Kmax}\}$	$\{\text{det}1,\text{det}2,....,\text{det}_{Kmax}\}$
		......	......	......	......
......	....	......	......	......	......

top-K query directly in mobile IoT, such as the original data block division issue, and the security issue of G_m and $y(e)$. We will deal with these issues and make NC applicable to top-K query.

C. SCHEME OVERVIEW

In this paper, we adopt three-party framework without any trusted entity, i.e., "Vehicle—Intermediate Node—SP" in Fig.3, where SP is honest-but-curious, i.e., SP strictly carries out query schemes and provides service to vehicles, but hopes to dig out more information for commercial interests. And all vehicles are not trusted by SP, either. SP wouldn't like to release other PoI record to vehicles. Our query scheme includes three phases and is briefly described as follows:

- An oblivious transfer query is generated by a vehicle and then sent to SP for the symmetric key to its desired PoI without revealing vehicle privacy or extra data of SP, as described in phase ① ② of Fig.3;
- Then, the vehicle initiates a PIR query to get the top-K PoIs coordinates and their details which are encrypted by the symmetric key mentioned above, and the SP cannot find any clue of the vehicle's interest or location in this phase, as described in phase ③④ of Fig.3;
- Finally, a vehicle generates an encrypted packet of previous query interest to a vehicle who is elected by other nearby vehicles, the elected vehicle performs encoding operation on the packets and delivers the combinations to the SP in different routes, the SP cannot decode the original interests until it receives enough combinations, which is essential to our top-K query, as described in phase ⑤⑥ of Fig.3.

D. A BENEFICIAL DATA STRUCTURE

We will introduce a beneficial data structure to store vehicle interests at SP side, such as the PoIs they queried. Assume that all PoIs in the map are distributed on distinct road segments in Fig. 4, we can similarly denote a vehicle u_k locating at a directed road segment $\overrightarrow{v_m v_n}$ as $u_k \in \overrightarrow{v_m v_n}$, and a vehicle initiates a query with its location and interest for top-K PoI set $P^K_{u_k}$,which includes two parts $P^K_{uk} = P^X_{uk} + P^{(K-X)}_{vn}$, where $P^X_{u_k}$ indicates a PoI set from u_k to v_n in front of the vehicle, and the same as the PoI set $P^{(K-X)}_{vn}$. If u_k wants to get to its desired PoI on any other road segment, it has to reach the vertex in front firstly anyway.

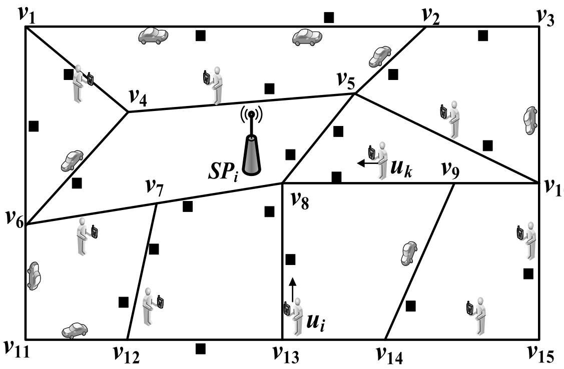


FIGURE 4. Description of the road networks.

Note that each vertex in front of a vehicle can be taken as a starting point to reach the PoIs nearby, so all the vertexes can be taken as generators to organize top-K PoI records in the data structure in TABLE 2. In another word, each vehicle initiates a query with its forward vertex as a dummy to obtain a top-K PoI set $P^K_{u_k}$, in which K PoIs are mainly ranked based on road distance, rather than the Euclidean distance which is adopted by most other KNN query schemes and leads to service inaccuracy because of the building blockage or road constrains. More than that, only one new query is generated by the vehicle as long as it enters into a new road segment $\overrightarrow{v_m v_n}$ with its forward vertex in mobile IoT, which significantly reduces the query number.

Six columns exist in TABLE 2 and in sequence indicate the name of vertex v_i in road networks, the neighbor vertexes of v_i, various PoI types departing from v_i, the unique code j of distinct PoI type, the coordinate set of a PoI type ranked by the following Equation (2) departing from vertex v_i and detailed information of different PoIs. Each detail of PoI records in the last two columns is encrypted with different symmetric keys in a bid to protect data security stored in SP. Accordingly, if a geography map contains m vertexes and at most n PoI types for each vertex, as shown in TABLE 2. the SP has to generate $m \times n$ symmetric keys $k_{ij}, (1 \leq i \leq m, 1 \leq j \leq n)$ for each PoI type. In addition, SP extracts the columns from 1 to 4 in TABLE 2 as an index and periodically releases it to the vehicles, so that the smart components in a vehicle can determine its road segment.

Furthermore, we consider more factors to rank the popularity of a PoI rather than only the distance in road networks, so we add a criterion value T_v to each PoI, T_v is computed by SP in the light of other factors, such as query rate, user reviews, financial profit, etc., such that the top-K PoIs can be calculated precisely by the following weighted Equation (2).

$$rank(P_i) = \alpha \cdot dis_{norm}(loc_{uk}, loc_{pi}) + (1-\alpha) \cdot T_v, \qquad (2)$$

wherein dis_{norm} is road distance function by means of taking the longest road segment in entire map as a road distance normalization factor, and $0 < \alpha < 1$ is the weighing factor.

In our scheme, to preserve the vehicle's privacy and SP's data security, the vehicle initially carries out the oblivious transfer protocol to insensibly acquire the unique symmetric key k_{ij} of its desired PoI in TABLE 2 and without revealing any other PoI record stored in the SP or the location and interest privacy of the vehicle. Then, the vehicle carries out the PIR protocol to get the top-K PoI coordinates and top-K detailed information in the last two columns of TABLE 2, which are encrypted by the symmetric key k_{ij}, whereas SP hardly relates the PoI record to any vehicle interest which already had been retrieved by a vehicle.

Finally, each vehicle encrypts its interest with SP's public key and transmit it to a sink vehicle who is elected to combinate their input packets, and the sink vehicle carries out a random linear encoding operation on these cryptographic packets and delivers them to SP. The SP decodes and decrypts the original vehicle interests after receiving enough combinations without relating these interests to any specific vehicle and then rearranges the PoI records in TABLE 2. Note that three protocols can be performed simultaneously.

We also assume that the whole map of the road networks is divided into different governing areas, each distributed SP server responds to query requests from vehicles and possess various information in its governing area, such as the map, the PoI distribution information in its own TABLE 2. As shown in Fig. 4, it's a governing area with a distributed SP$_i$.

III. OUR SCHEME

As mentioned before, SP generates distinct keys, such as a key k_{ij}, to separately encrypt the top-K coordinates and details of each PoI record. Owing to the correlation between the symmetric key and a PoI, the vehicle privacy will be exposed if the vehicle queries for the key k_{ij} directly. To deal with this issue, we utilize oblivious transfer to obtain a symmetric k_{ij} from the SP privately.

Originally, SP encrypts each key k_{ij} by Equation (3), wherein $H(\cdot)$ and $\oplus$ indicate a hash function and an exclusive or operation respectively in the equation, we set $keyij = g^{a_i}||g^{b_j}, (1 \leq i \leq m, 1 \leq j \leq n)$, wherein a_i and b_j are integers randomly picked by SP and correspond to different rows and columns in TABLE 2. Each row of the same vertex v_i has the same a_i and distinct b_j, which reduces the computational overhead than other solutions without security compromise, the symbol || is a concatenating operation. Then a vehicle can obtain a unique key K_{ij} each time by the oblivious transfer protocol in the following subsection III.A.

$$K_{ij} = k_{ij} \oplus H(key_{ij})g \qquad (3)$$

In the light of the beneficial data structure in the subsection II.D, we are able to implement the following two cryptographic protocols more efficiently in vehicle networks.

A. OUR OT PROTOCOL

Aiming at mobile IoT scenario in road networks, we propose an oblivious transfer (OT) protocol based on Elgamal cryptosystem [35], which works between vehicles and SP in our query scheme with SP data and vehicle privacy preservation. We will firstly describe the OT protocol in this subsection and its interaction process between a vehicle and the SP is show in TABLE 3, it contains 3 steps:

Step 1: Owing to travel restrictions from building blockage or road constraint, a vehicle u_k has to travel along the road segment until it reaches the forward vertex and enters a new segment to get to desired PoIs. Thus, the vehicle adopts its forward vertex v_i as a starting point in a query rather than its actual location to retrieve top-K PoI records, and then it utilizes the index table from SP to determine the unique code j of its desired PoI record from vertex v_i. Besides, the i and j reveal the vehicle's location and interest privacy when it queries for the key *keyij*, i.e., the vehicle locates on a road segment with vertex v_i and it queries for the jth PoI type in TABLE 2. To deal with this issue, we utilize oblivious transfer to encrypt the i in symbol v_i and the unique code j, then a vehicle sends the query $(\boldsymbol{OT}_1, \boldsymbol{OT}_2)$ to SP for *keyij*.

Step 2: In order to acquire the corresponding key secretly from SP, as soon as receiving the OT query $(\boldsymbol{OT}_1, \boldsymbol{OT}_2)$, for each integer ξ and η within $1 \leq \xi \leq m$ and $1 \leq \eta \leq n$, SP calculates two candidate vectors $(\boldsymbol{R}, \boldsymbol{S})$ as potential query result for vehicle to calculate the key *keyij*, then the cryptographic vector $(\boldsymbol{R}, \boldsymbol{S})$ are transmitted to the vehicle.

Step 3: On the basis of oblivious transfer, a vehicle is merely capable to calculate its desired key key_{ij} according to $(\boldsymbol{OT}_{1,i}, \boldsymbol{OT}_{2,j})$ in $\boldsymbol{R}$ and $\boldsymbol{S}$ when $\xi = i, \eta = j$, and it computes $key_{ij} = w_1||w_2$ by the following Equation (4):

$$\begin{aligned} w_1||w_2 &= Y_{1,i}/(X_{1,i})^x||Y_{2,j}/(X_{2,j})^x \\ &= g^{a_i}(g^i\delta_1)^{r_i}/(\varepsilon_1^{r_i})^x||(g^{b_j}(g^j\delta_2)^{r_j}/\varepsilon_2^{r_j})^x \\ &= [g^{a_i}(g^i g^{-i}(g^x)^{r_1})^{r_i}]/(g^{r_1 r_i})^x|| \\ &\quad \times [g^{b_j}(g^j g^{-j}(g^x)^{r_2})^{r_j}]/(g^{r_2 r_j})^x \\ &= g_1^{a_i}||g_2^{b_j} \end{aligned} \qquad (4)$$

and then it further calculates the $k_{ij} = K_{ij} \oplus H(key_{ij})$ in the light of Equation (3) by its self-inverse of $\oplus$ operator. With

TABLE 3. The interaction process of our OT protocol between a vehicle and SP.

Our OT Protocol	
Vehicle	**Service Provider (SP)**
	$\longleftarrow$ release an index
determine user's desired *j*th PoI of the *i*th vertex	
generate $pk = g^x$, x is random integer	
calculate $\boldsymbol{OT}_1 = E(i, pk) = (\varepsilon_1, \delta_1) = (g^{r_1}, g^{-i} pk^{r_1})$	
calculate $\boldsymbol{OT}_2 = E(j, pk) = (\varepsilon_2, \delta_2) = (g^{r_2}, g^{-j} pk^{r_2})$	
initiate an OT query $\xrightarrow{(\boldsymbol{OT}_1,\boldsymbol{OT}_2)}$	
	for each $1 \le \xi \le m \,\&\, 1 \le \eta \le n$
	calculate $\boldsymbol{OT}_{1,\xi} = (\varepsilon_1^{r_\xi}, g^{a_\xi}(g^{\xi}\delta_1)^{r_\xi})$,
	calculate $\boldsymbol{OT}_{2,\eta} = (\varepsilon_2^{r_\eta}, g^{b_\eta}(g^{\eta}\delta_2)^{r_\eta})$
	let $\boldsymbol{R} = (\boldsymbol{OT}_{1,1},..., \boldsymbol{OT}_{1,\xi},..., \boldsymbol{OT}_{1,m})$ and $\boldsymbol{S} = (\boldsymbol{OT}_{2,1},..., \boldsymbol{OT}_{2,\eta},..., \boldsymbol{OT}_{2,n})$
	$\xleftarrow{(\boldsymbol{R},\boldsymbol{S})}$ return a cryptographic result
pick $(X_{1,i}, Y_{1,i}) = \boldsymbol{OT}_{1,i}$, $(X_{2,j}, Y_{2,j}) = \boldsymbol{OT}_{2,j}$	
calculate $w_1 = Y_{1,i} / (X_{1,i})^x$, $w_2 = Y_{2,j} / (X_{2,j})^x$	
calculate $key_{ij} = w_1 \parallel w_2$	
return $k_{ij} = K_{ij} \oplus H(key_{ij})$	

the symmetric key, the vehicle is capable to only decrypt one PoI detail it desired before.

B. OUR PIR PROTOCOL

Similarly, we devise a PIR protocol based on Gentry et al's cryptosystem [28] and it works between vehicles and SP. The novel improvements in this paper are these: we implement a PIR protocol in the real road networks scenario rather than in the Euclidean space, which makes the query result more accurate than other typical solutions [3], [4], [28]. And more than that, our PIR protocol is capable to acquire the desired top-K PoIs accurately just at one time, rather than excessively searching for the K PoIs in broad area around the vehicle successively in other similar solutions [3], [4], [28], our PIR protocol significantly decreases the query scopes and communication cost in this paper.

Besides, in mobile IoT scenario, the vehicle sends a query only once as soon as it gets in another segment rather than issuing the same query to SP periodically according to location updates in the solutions [3], [4]. Its interaction process of our PIR protocol between a vehicle and the SP is show in TABLE 4, it contains 3 steps.

In our PIR protocol, we assume all PoI records belonging to a certain vertex v_i in terms of a data block, denoted as an integer C_{vi}. Then an integer set $PE = \{p_1^{N_1}, p_2^{N_2}, \ldots, p_m^{N_m}\}$ are calculated by SP and assigned to each data block corresponding to distinct vertexes, wherein $\omega_i = p_i^{Ni}$ in PE is a prime power, p_i and N_i are respectively a prime number and a relatively small natural number. Particularly, the size of each data block is smaller than its relevant p_i^{Ni}, such that $|C_{vi}| < \omega_i$, all vehicles are aware of the set PE. At last, a minimum integer C_{LBS} is calculated with Chinese Remainder Theorem, such that $C_{LBS} = C_{vi}(\mathrm{mod}\,\omega_i)$, $1 \le i \le m$, wherein the C_{LBS} indicates the database of a certain distributed SP who governs an area in a divided map.

In our PIR protocol, a vehicle initially determines the vertex v_i in front of its current road segment, and its relevant $\omega_i = p_i^{N_i}$ which divides $| < g > |$, the order of an element g in the random group G generated by vehicle. Eventually, the vector (G, g) is sent back to SP.

After receiving (G, g), SP computes $g^{C_{LBS}}$ straightway and delivers it to the vehicle. Then, the vehicle is capable to calculate desired data block C_{vi} by following Equation (5).

$$\begin{aligned} C_{vi} &= \log_w(g^{\mu c}) = \log_w(g^{c|<g>|/\omega_i}) = \log_w(g^{c_{\omega_i}|<g>|/\omega_i}) \\ &= \log_w((g^{|<g>|/\omega_i})^{c_{\omega_i}}) \\ &= \log_w((g^{\mu})^{c_{\omega_i}}) = \log_w(w^{c_{\omega_i}}) = c_{\omega_i} \end{aligned} \tag{5}$$

wherein c_{ω_i} indicates the calculation $C_{LBS}(\bmod \omega_i)$. With the acquired symmetric key before, the vehicle can uniquely decrypt one PoI record desired in the data block C_{vi} without revealing any vehicle privacy to SP, or extra PoI data in SP to

TABLE 4. The interaction process of our PIR protocol between a vehicle and SP.

Our PIR Protocol	
Vehicle	**Service Provider (SP)**
determine the vertex v_i in front of him and the relevant $\omega_i = p_i^{N_i}$ generate a group G with an element g the order $\lvert < g > \rvert$ can be divided by ω_i send the vector $\xrightarrow{(G,g)}$	
	calculate $g^{C_{LBS}}$ $\xleftarrow{g^{C_{LBS}}}$ return the cryptographic result
calculate $\mu = \lvert < g > \rvert / \omega_i$, $w = g^{\mu}$ calculate and return the result $C_{vi} = \log_w(g^{\mu C_{LBS}})$	

vehicle. Therefore, the vehicle privacy and SP data security are well protected in this PIR scheme.

C. IMPLEMENTATION OF THE TOP-K QUERY

The two above-mentioned protocols aim at getting K PoIs without revealing vehicle privacy or SP data. Such query scheme, however, cannot achieve top-K query since SP has to know the query interests from vehicles to rearrange the top-K PoI records in its database, which is incompatible to the core idea of PIR. We will maintain PIR with achieving top-K query by adding networks coding (NC) into our query scheme. With the cooperation of intelligent vehicles in mobile IoT, the SP cannot calculate the query interest from vehicles until enough NC packets are collected, and it cannot relate any of these query interests to a specific vehicle as well, which insures PIR against being invalid.

In the road networks, each vehicle in different road segments can be considered as a member in a vehicle ad hoc network. After or before a query, a vehicle encrypts its query interest with the public key published by the SP, and delivers the ciphertext to a nearby vehicle who is elected randomly by its neighbor vehicles as a cluster head (receiver), the election algorithms [39], [40] and relevant data protections [41]–[44] have been well explored in previous work, we will not focus on it in this paper. The vehicle changes its pseudonym in each snapshot query to prevent correlations.

The algebraic relations mentioned in the subsection II.B can be applied to these query interest packets. After receiving h packets $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_h)$, the receiver adds a unit matrix to these packets, which fundamentally helps packets to record each linear transformation on intermediate nodes and facilitates the decoding operations at SP, denoted as the following Equation (6):

$$\boldsymbol{X} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} & 1 & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2k} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{h1} & a_{h2} & \cdots & a_{hk} & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{6}$$

and then the receiver generates its local encoding coefficient matrix and calculates a combination of these packets by Equation (7), wherein $\boldsymbol{\beta}_i$ is the ith line of matrix $\boldsymbol{X}$. The h packets can be accordingly called as the corresponding packets of aforementioned combination.

$$\begin{aligned} &\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1h} \\ c_{21} & c_{22} & \cdots & c_{2h} \\ \vdots & \vdots & \vdots & \vdots \\ c_{h1} & c_{h2} & \cdots & c_{hh} \end{bmatrix} \boldsymbol{X} \\ &\quad = \begin{bmatrix} c_{11}\boldsymbol{\beta}_1 + c_{12}\boldsymbol{\beta}_2 + \ldots + c_{1h}\boldsymbol{\beta}_h \\ c_{21}\boldsymbol{\beta}_1 + c_{22}\boldsymbol{\beta}_2 + \ldots + c_{2h}\boldsymbol{\beta}_h \\ \vdots \\ c_{h1}\boldsymbol{\beta}_1 + c_{h2}\boldsymbol{\beta}_2 + \ldots + c_{hh}\boldsymbol{\beta}_h \end{bmatrix} \end{aligned} \tag{7}$$

Finally, the receiver delivers the combination and corresponding packets with different unit vectors in h distinct paths to SP.

Actually, the network is dynamic in mobile IoT, with vehicle login or logout in an ad hoc way. But these query combinations and their corresponding packets are always forwarded to different sink nodes, such as roadside units, base stations, etc. After receiving a certain number of the packets of the same combination from a receiver, a sink node will recode the packets with its own local encoding vector by performing random linear coding and output a new combination, so each output can be taken as a combination of original h packets $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_h)$ in Equation (8):

$$\begin{aligned} [\boldsymbol{y}(e), \boldsymbol{g}(e)] &= \sum\nolimits_{e'} C_{e'}(e)[\boldsymbol{y}(e'), \boldsymbol{g}(e')] \\ &= \sum\nolimits_{i=1}^{h} g_i(e)[\boldsymbol{E}_i, \boldsymbol{x}_i] \end{aligned} \tag{8}$$

where $\boldsymbol{E}_i$ is a unit matrix, and each globe encoding coefficient $\boldsymbol{g}(e)$ is time varying and diversified, but only the terminal node, i.e., SP, is capable to decode the combinations due to adding the unit vectors which records the random linear transformation during the transmission and facilitates the calculation of $\mathbf{G}_m$ at SP without being aware of the network topology knowledge or message routing path.

For sake of the confidentiality of $\mathbf{G}_m$, homomorphic encryption with the public key from SP can be implemented the on original encoding coefficient at each selected receiver, and at each sink node the linear recoding can be carried out on the encrypted coefficient which is equivalent to the same operation on the plaintext due to the homomorphic feature [28], [31], so any other malicious attacker in the vehicle networks cannot obtain the $\mathbf{G}_m$, so the original interest packets of vehicles are protected well in the transmission.

After gathering enough combining packets of the same generation and decrypting the coefficient $\mathbf{G}_m$, SP is capable to decode the original vehicle interest with the following Equation (9) as long as $\mathbf{G}_m$ is invertible.

$$\begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_h \end{bmatrix} = \mathbf{G}_m^{-1} \begin{bmatrix} \boldsymbol{y}(e_1) \\ \vdots \\ \boldsymbol{y}(e_h) \end{bmatrix} \tag{9}$$

The invertibility of the matrix $\mathbf{G}_m$ can be easily achieved as long as the corresponding packets at a receiver are forwarded in distinct edge-separating paths [34]–[37]. It can be seen that each combination seems like the same as an encryption and meaningless to any malicious attacker, and only after gathering enough combinations, SP can calculate the original packets which are the exact query interests from cooperative vehicles, and then SP can periodically update its top-K PoI records stored in TABLE 2 without inferring any of these query interests to a specific vehicle. To sum up, the aforementioned three phases, OT, PIR and NC constitute a scheme that achieves the top-K query with vehicle's location and query interest privacy preservation and SP's data security guarantees in mobile IoT scenario.

IV. PERFORMANCE ANALYSIS

A. PRIVACY AND SECURITY ANALYSIS

In the our oblivious transfer protocol, only one symmetric key can be acquired by the vehicle each time, where the encryption key $key_{ij} = w_1||w_2 = g^{a_i}||g^{b_j}$ can be worked out if and only if $\xi = i$ and $\eta = j$ in Equation (4) and the following Equation (10).

$$\begin{aligned} w_1||w_2 &= g^{a_\xi}(g^i\delta_1)^{r_i}/(\varepsilon_1^{r_i})^x||(g^{b_j}(g^j\delta_2)^{r_j}/\varepsilon_2^{r_j})^x \\ &= [g^{a_\xi}(g^i g^{-\xi}(g^x)^{r_1})^{r_\xi}]/(g^{r_1 r_\xi})^x|| \\ &\quad \times [g^{b_\eta}(g^j g^{-\eta}(g^x)^{r_2})^{r_\eta}]/(g^{r_2 r_\eta})^x \\ &= g^{a_\xi}g^{i-\xi}||g^{a_\eta}g^{j-\eta} \end{aligned} \tag{10}$$

Because only the vehicle knows x, and it is hard to work out other keys in term of brute force due to the intractability of discrete logarithm, and the vehicle cannot deduce any other encryption key of vertex v_i with the same g^{a_i}, either.

Respectively, the i of v_i and the unique code j imply location privacy and query interest privacy of a vehicle, we utilize ElGamal [38] to encrypt them, such that SP or attackers are hard to crack vehicle's relevant privacy on account of the semantic security provided by ElGamal. During the whole process of our scheme, the vehicle's locations are generalized uniformly to the entire governing area of a certain SP. Compared with the schemes in [3], [15], in which a vehicle is restricted to a small region, it's easy for the vehicle to be identified by a malicious attacker if the vehicle is the only one in that region.

TABLE 5. Computational cost in OT protocol.

	Vehicle	SP	Total
Reference [3]	7	$3(m+n)+1$	$\gamma[3(m+n)+8]$
Reference [15]	$4mn+4$	$4mn$	$\gamma(8mn+4)$
Our scheme	6	$3(m+n)$	$3(m+n)+6$

Additionally, SP is never aware of any vehicle interest, e.g., the $g^{C_{LBS}}$ in PIR protocol is calculated by SP who gets no significant information rather than an exponentiation, and consequently the vehicle works out its desired top-K PoIs itself with $g^{C_{LBS}}$. Despite many similar solutions have already made great improvements [3], [15], we achieve preferable accuracy and efficiency with costly cryptography for mobile IoT in real road networks.

Cooperative vehicles' interests are also preserved well in the NC procedure since each packet is a combination of h vehicle query interests in Equation (8), due to none of these combinations signifies an explicit query, the SP is hard to relate it to any specific vehicle. Moreover, homomorphic encryption is employed on original local encoding coefficient so that each linear recoding can be confidentially carried out on each sink nodes, which effectively guarantees the query interest privacy in mobile IoT.

Besides, traffic analysis can be performed on intermediate nodes or malicious attackers in the NC phase. Generally, there are three typical technologies in traffic analysis: size correlation, time order correlation and content correlation. In our scheme, each packet or combination is set to the same size to resist the size correlation, and it's hard to implement time order correlation since each packet or combination can be randomly sent to the SP. We also assume the resistances to the aforementioned size or time order correlation are compromised, the original global encoding coefficient can be encrypted by Paillier cryptosystem, attackers are hard to decoded the combinations without being aware of the $\mathbf{G}_m$, even if an attacker compromises several intermediate nodes to analyze the query interests from vehicles.

Moreover, the sink vehicle is also elected randomly, the cost is extremely expensive for attackers to collude with each potential receiver, the vehicle can preserve its query interest well by means of querying with a new pseudonym each time even if SP colludes with a large number of vehicles. Furthermore, NC also helps to enhance transmission capability in vehicle networks without privacy compromise.

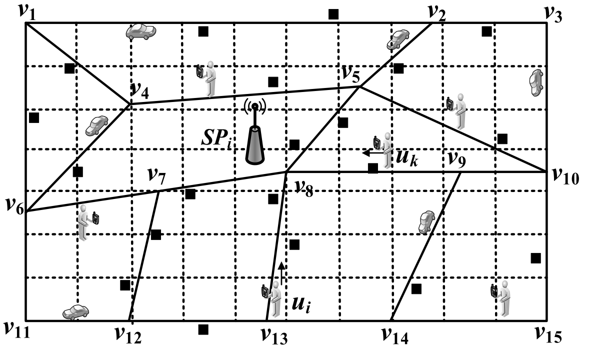


FIGURE 5. Divide the map into cells.

B. COMPUTATIONAL COST

As most researches concern, we will focus on the computation cost analysis and make a comparison between our work and similar solutions [3], [15].

1) OT PROTOCOL ANALYSIS

Similar to [3], [15], we take modular exponentiation as the major cost in our scheme. As shown in TABLE 3, a vehicle performs encrypting operation to generate an OT query in $\boldsymbol{OT}_1 = E(i, pk) = (g^{r_1}, g^{-i+xr_1})$, therefore 4 exponentiations are executed in $\boldsymbol{OT}_1$ and $\boldsymbol{OT}_2$. After receiving the cryptographic result, the vehicle performs $w_1 = Y_{1,i}\big/(X_{1,i})^x$ to calculate the accurate query result, therefore 2 exponentiations are executed in calculating w_1 and w_2. Thus, 6 exponentiations are executed in an OT query at the vehicle side in total, as shown in TABLE 5.

SP carries out 4 exponentiations in $(\varepsilon_1^{r_\xi}, g^{a_\xi}(g^\xi \delta_1)^{r_\xi})$ for every $\xi \in [1, m]$ to create an $\boldsymbol{OT}_{1,\xi}$ after receiving a query. And g^ξ can be calculated ahead of time for cutting down the processing time since SP possesses the element g, and consequently $3m$ exponentiations are executed for calculating the $(\boldsymbol{OT}_{1,1}, \ldots, \boldsymbol{OT}_{1,m})$, take another $3n$ exponentiations in $(\boldsymbol{OT}_{2,1}, \ldots, \boldsymbol{OT}_{2,n})$ into account, the number of exponentiations adds up to $3(m+n)$ at SP side, as shown in TABLE 5.

Finally, another two exponentiations are performed at vehicle side to calculate w_1, w_2. The computational costs in our protocol are compared with two typical schemes [3], [15] who employ the OT protocol in their scheme as well, the comparison of computational cost is shown in TABLE 5.

These solutions [3], [15] whose geographical maps are all described with cells in Fig. 5. For obtaining the latest PoI query result in mobile IoT, the vehicle has to issue a new query as long as it gets in a new cell accompanied with location updates. But in our scheme, the vehicle only issues one query as long as it gets into a new road segment which traverses a number of cells. That implies, if u_i enters into $\overrightarrow{v_{13}v_8}$ which traverses four cells, it only issues one query in our query scheme rather than issue four queries in [3], [15]. As a result, if each road segment traverses a certain of cells at an average of γ, the computing cost is $1/\gamma$ of solutions [3], [15].

2) PIR PROTOCOL ANALYSIS

In [15], PIR query is devised in Euclidean space with $m \times n$ cells, and accordingly the number of multiplication is $N\sqrt{m \times n}$ and $m \times n$ executed at vehicle and SP ends in TABLE 6, respectively. Compared with that, the scheme in [3] contains 3 exponentiations, and the number of its multiplication is $2|N|$ and $|e|$ at vehicle and SP end. Four exponentiations exist in our scheme for homomorphic encryption on encoding coefficient at the sink nodes, $(S_{NC} + 1) \times h$ exponentiations in recoding process, where S_{NC} and h are the average number of sink nodes and the combinations received on each node, respectively. And owing to utilizing the Pohlig-Hellman algorithm, it brings in a process with a complexity of $O(N \log p^N + \sqrt{p})$ in [3], whereas we have a process with a complexity of $O(\sqrt{p})$ in the logarithmic solving, which improves the efficiency.

Thus, the PIR computational overhead of three mentioned schemes is summarized in TABLE 6, where *exp* represents the cost of an exponentiation. It shows that our scheme has an advantage of computational overhead in a snapshot query. More than that, if each road segment traverses γ cells on average, the computational cost of our mobile service request is $1/\gamma$ than the other two schemes in road networks. We do not consider the computational cost in NC due to its linear operations on sink nodes in mobile IoT scenario, which are negligible compared with modular exponentiations.

Furthermore, we adopt distributed SP whose service data set in TABLE 2 is smaller than that in a centralized database, so when the minimum integer $C_{LBS} = C_{vi}(\text{mod}\omega_i)$, $1 \leq i \leq m$ is calculated by SP with the Chinese Remainder Theorem to generate a cryptographic result, in which C_{LBS} indicates the database of a certain distributed SP who governs an area in a divided map, the computational overhead is obviously much less than that in the centralized database. So due to the less computation amount, our scheme generates query result faster than the other solutions who are also built on strong cryptology, it is applicable to the mobile IoT scenario.

3) NETWORK CODING ANALYSIS

The computational overhead can be investigated from three aspects: original encoding on the select cluster vehicle, recoding on the intermediate node and decoding on the SP.

Consider an $h \times h$ encoding matrix G_m on the cluster vehicle, for sake of security and recoding on intermediate nodes, each element in G_m can be encrypted by Paillier cryptosystem to make good use of its homomorphic property. So that the encryption computational cost on G_m is $O(h^2)$ according to Equation (7), and each encrypting operation includes 1 multiplication, 1 modulus and 2 exponentiations. Thus, the computation cost is $O(h^2 \cdot \log n)$ in total.

Intermediate nodes can perform random linear recoding on the ciphertext of inputting encoding packets, which is equivalent to the recoding operation on plaintext packets.

TABLE 6. Computational cost in PIR protocol.

	Vehicle	SP	Total
Reference [3]	$O(N\log p^{N}+\sqrt{p})+2\|N\|+2exp$	$\|e\|+exp$	$\gamma[O(N\log p^{N}+\sqrt{p})+2\|N\|+3exp+\|e\|]$
Reference [15]	$N\sqrt{m\times n}$	$m\times n$	$\gamma(N\sqrt{m\times n}+m\times n)$
Our scheme	$O(\sqrt{p})+2\|N\|+S_{NC}\times h+4exp$	$\|e\|+h+3exp$	$O(\sqrt{p})+2\|N\|+\|e\|+(S_{NC}+1)\times h+7exp$

TABLE 7. Configurations.

Parameters	Range of value	Default
Number of vehicles/ users U	$5\leq U\leq 30$	15 ($\times 10^4$)
Number of PoIs K	$5\leq K\leq 25$	10
Number of PoIs m	$2.5\leq m\leq 10$	10 ($\times 10^4$)
Average length of road S	$200\leq S\leq 2000$	1000 meters
Location updating frequency f	$5\leq f\leq 30$	15 seconds

Similarly, the computational cost on each local vector $\boldsymbol{g}(e)$ is $O(h^2\cdot\log n)$. According to the Equation (8), the total computational cost of a G_m with $h\boldsymbol{g}(e)$ is $O(h^3\cdot\log n)$ due to the multiplication operation provided by the homomorphic property of Paillier cryptosystem.

Only SP can decrypt the encoded combination according to Paillier cryptosystem, it includes 1 multiplication, 1 exponentiation and 1 division operation on a combination, so the computational cost is $O(h\cdot\log n)$. Only after collecting h linear independent combinations, SP can decode the original packets, so the computational cost is $O(h^3)$.

V. EXPERIMENTS

We employ the geographical data in dataset OL [45] and generate mobile objects with TBG [46] as vehicle continuous locations in road networks. Our scheme was implemented on Windows 7 platform by Java programs. The primary parameters are recorded as follows in TABLE 7, each service request packet is normally set to 1500 bytes in total, in which includes $(1500-40)/300\approx 5$ PoI records as defined that each PoI record and the packet header cost 300 bytes and 40 bytes, respectively.

A. QUERY ACCURACY

Vehicles are capable to obtain accurate top-K PoIs according to its interest, which are arranged and stored in advance in the data structure of TABLE 2, it implies the top-K PoIs are the nearest ones along road segments without blockages in mobile IoT scenario rather than the Euclidean space, such as building blockage, road restrictions. All the top-K records are rearranged corresponding to the updating popularity of vehicle interests by introducing NC. The query accuracy is compared with [3], [15] in Fig. 6 and Fig. 7.

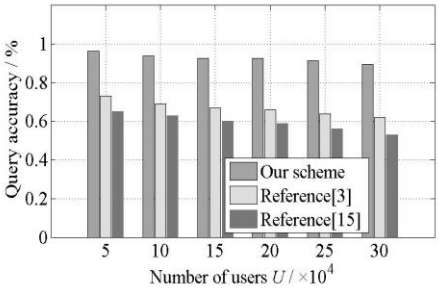


FIGURE 6. Comparison of query accuracy (U varies).

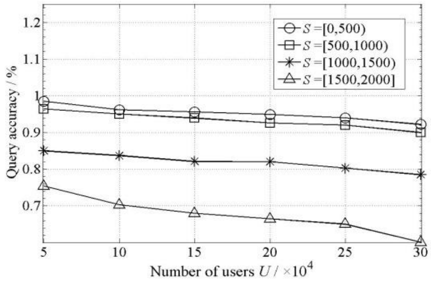


FIGURE 7. Comparison of query accuracy (S, U vary).

As show in Fig.6, our scheme is more accurate than [3], [15] since it is devised based on real road networks, e.g., despite the KNN PoIs are gotten in [3], [15], these PoIs are distributed in the Euclidean space, which may be unreachable for a long time along the road, so it implies the accuracy for querying PoIs cannot be guaranteed all the time. As shown in Fig. 7, when the average length of road segments S becomes longer, the query accuracy decreases as a result of discarding some PoIs caused by pruning and merging in our scheme.

B. PROCESSING TIME AND COMMUNICATIONS

As shown in Fig.8, when the number of vehicles increasing, the SP faces more service requests and hence the average processing time of three schemes increases to a certain extent.

Our scheme has lower computational cost compared with the scheme [3]. We also have a certain of preponderances on account of query number declines and discrete logarithm improvements in PIR phase, its average processing time reaches the half of [15], whereas a mobile user in [3], [15] has to send more queries in different cells.

Our scheme makes expensive cryptographic scheme more practical due to the reduction of query number. As shown in Fig. 9, as location updating in time interval from 30s to 5s, the query packets tend to increase rapidly in [3], [15], whereas

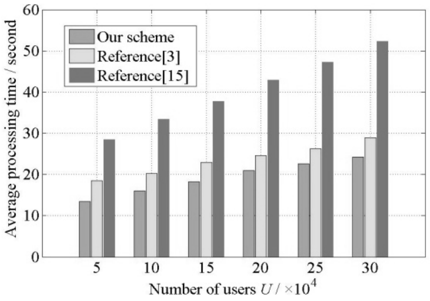


FIGURE 8. Average processing time (U varies).

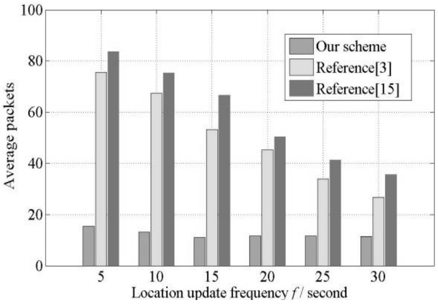


FIGURE 9. Average packets (f varies).

our scheme remains stable because the vehicle doesn't issue queries corresponding to the location updating.

VI. CONCLUSION

To preserve the private data which are collected by the smart components in vehicle networks, we propose a top-K query scheme with privacy preservation by introducing NC into mobile IoT scenario for the first time. In our scheme, vehicles with smart components cooperate with each other to access to mobile service without revealing their location or interest privacy, and our scheme has more accurate service result since taking full consideration of the factors in real road network. Moreover, it has less computational and communicational overhead even built on strong cryptology. Compared with existing solutions, we verified its accuracy and efficiency in performance analysis and experiments.

REFERENCES

[1] X. Zeng, G. Xu, X. Zheng, Y. Xiang, and W. Zhou, "E-AUA: An efficient anonymous user authentication protocol for mobile IoT," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1506–1519, Apr. 2019.

[2] P. Asuquo, H. Cruickshank, J. Morley, C. P. A. Ogah, A. Lei, W. Hathal, S. Bao, and Z. Sun, "Security and privacy in location-based services for vehicular and mobile communications: An overview, challenges, and countermeasures," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4778–4802, Dec. 2018.

[3] R. Paulet, M. G. Kaosar, X. Yi, and E. Bertino, "Privacy-preserving and content-protecting location based queries," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 5, pp. 1200–1210, May 2014.

[4] X. Yi, R. Paulet, E. Bertino, and V. Varadharajan, "Practical approximate k nearest neighbor queries with location and query privacy," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 6, pp. 1546–1559, Jun. 2016.

[5] T. Ma, J. Jia, Y. Xue, Y. Tian, A. Al-Dhelaan, and M. Al-Rodhaan, "Protection of location privacy for moving kNN queries in social networks," *Appl. Soft Comput.*, vol. 66, pp. 525–532, May 2018.

[6] M. Elkhodr, S. Shahrestani, and H. Cheung, "A review of mobile location privacy in the Internet of Things," in *Proc. 10th Int. Conf. ICT Knowl. Eng.*, Nov. 2012, pp. 266–272.

[7] J. Zhou, Z. Cao, X. Dong, and A. V. Vasilakos, "Security and privacy for cloud-based IoT: Challenges," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 26–33, Jan. 2017.

[8] L. Chen, S. Thombre, K. Jarvinen, E. S. Lohan, A. Alen-Savikko, H. Leppakoski, M. Z. H. Bhuiyan, S. Bu-Pasha, G. N. Ferrara, S. Honkala, J. Lindqvist, L. Ruotsalainen, P. Korpisaari, and H. Kuusniemi, "Robustness, security and privacy in location-based services for future IoT: A survey," *IEEE Access*, vol. 5, pp. 8956–8977, 2017.

[9] X. Pan, J. Xu, and X. Meng, "Protecting location privacy against location-dependent attacks in mobile services," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 8, pp. 1506–1519, Aug. 2012.

[10] P. Galdames, C. Gutierrez-Soto, and A. Curiel, "Batching location cloaking techniques for location privacy and safety protection," *Mobile Inf. Syst.*, vol. 2019, Jan. 2019, Art. no. 9086062.

[11] C. Zhou, T. Wang, W. Jiang, and H. Tian, "Practical k nearest neighbor query scheme with two-party guarantees in road networks," in *Proc. 17th IEEE Int. Conf. Trust, Secur. Privacy Comput., 12th IEEE Int. Conf. Big Data Sci. Eng. (TrustCom/BigDataSE)*, New York, NY, USA, Aug. 2018, pp. 1316–1321.

[12] M. L. Yiu, C. S. Jensen, X. Huang, and H. Lu, "SpaceTwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services," in *Proc. IEEE 24th Int. Conf. Data Eng.*, Apr. 2008, pp. 366–375.

[13] Z. Gong, G.-Z. Sun, and X. Xie, "Protecting privacy in location-based services using K-anonymity without cloaked region," in *Proc. 11th Int. Conf. Mobile Data Manage.*, Kansas City, MO, USA, 2010, pp. 366–371.

[14] Y. Huang, Z. Huo, and X.-F. Meng, "CoPrivacy: A collaborative location privacy-preserving method without cloaking region," *Chin. J. Comput.*, vol. 34, no. 10, pp. 1976–1985, Oct. 2011.

[15] G. Ghinita, P. Kalnis, M. Kantarcioglu, and E. Bertino, "Approximate and exact hybrid algorithms for private nearest-neighbor queries with database protection," *GeoInformatica*, vol. 15, no. 4, pp. 699–726, Oct. 2011.

[16] T. Wang, M. Z. A. Bhuiyan, G. Wang, L. Qi, J. Wu, and T. Hayajneh, "Preserving balance between privacy and data integrity in edge-assisted Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 2679–2689, Apr. 2020.

[17] H. Tao, M. Z. A. Bhuiyan, M. A. Rahman, T. Wang, J. Wu, S. Q. Salih, Y. Li, and T. Hayajneh, "TrustData: Trustworthy and secured data collection for event detection in industrial cyber-physical system," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3311–3321, May 2020.

[18] T. Wang, P. Wang, S. Cai, Y. Ma, A. Liu, and M. Xie, "A unified trustworthy environment establishment based on edge computing in industrial IoT," *IEEE Trans. Ind. Informat.*, to be published.

[19] T. Wang, H. Luo, X. Zheng, and M. Xie, "Crowdsourcing mechanism for trust evaluation in CPCS based on intelligent mobile edge computing," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 6, pp. 1–19, Dec. 2019.

[20] X. Li, K. Xie, X. Wang, G. Xie, D. Xie, Z. Li, J. Wen, and Z. Diao, "Quick and accurate falsedata detection in mobile crowd sensing," *IEEE/ACM Trans. Netw.*, early access, Apr. 2020, doi: 10.1109/TNET.2020.2982685.

[21] T. Wang, Z. Cao, S. Wang, J. Wang, L. Qi, A. Liu, M. Xie, and X. Li, "Privacy-enhanced data collection based on deep learning for Internet of vehicles," *IEEE Trans. Ind. Informat.*, early access, Dec. 2019, doi: 10.1109/TII.2019.2962844.

[22] T. Wang, D. Zhao, S. Cai, W. Jia, and A. Liu, "Bidirectional prediction-based underwater data collection protocol for end-edge-cloud orchestrated system," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4791–4799, Jul. 2020.

[23] T. Wang, L. Qiu, A. K. Sangaiah, A. Liu, M. Z. A. Bhuiyan, and Y. Ma, "Edge computing based trustworthy data collection model in the Internet of Things," *IEEE Internet Things J.*, early access, Jan. 2020, doi: 10.1109/JIOT.2020.2966870.

[24] Q. A. Arain, I. Memon, Z. Deng, M. H. Memon, F. A. Mangi, and A. Zubedi, "Location monitoring approach: Multiple mix-zones with location privacy protection based on traffic flow over road networks," *Multimedia Tools Appl.*, vol. 77, no. 5, pp. 5563–5607, Mar. 2018.

[25] Q. A. Arain, D. Zhongliang, I. Memon, S. Arain, F. K. Shaikh, A. Zubedi, M. A. Unar, A. Ashraf, and R. Shaikh, "Privacy preserving dynamic pseudonym-based multiple mix-zones authentication protocol over road networks," *Wireless Pers. Commun.*, vol. 95, no. 2, pp. 505–521, Jul. 2017.

[26] W. Eltarjaman, R. Dewri, and R. Thurimella, "Private retrieval of POI details in Top-K queries," *IEEE Trans. Mobile Comput.*, vol. 16, no. 9, pp. 2611–2624, Sep. 2017.

[27] L. Wang, R. Ma, and X. Meng, "Evaluating k nearest neighbor query on road networks with no information leakage," in *Proc. Int. Conf. Web Inf. Syst. Eng.*, Miami, FL, USA: Springer, 2015, pp. 508–521.

[28] C. Gentry and Z. Ramzan, "Single-database private information retrieval with constant communication rate," in *Proc. Int. Colloq. Automata, Lang., Program.*, Lisbon, Portugal, 2005, pp. 803–815.

[29] Z. Li, C. Ma, and D. Wang, "Towards multi-hop homomorphic identity-based proxy re-encryption via branching program," *IEEE Access*, vol. 5, pp. 16214–16228, 2017.

[30] Z. Li, C. Ma, and H. Zhou, "Multi-key FHE for multi-bit messages," *Sci. China Inf. Sci.*, vol. 61, no. 2, pp. 266–277, Feb. 2018.

[31] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proc. Int. Conf. Theory Appl. Cryptograph. Techn.* Prague, Czech Republic: Springer, 1999, pp. 223–238.

[32] X. Meng, H. Zhu, and G. Kollios, "Top-k query processing on encrypted databases with strong security guarantees," in *Proc. IEEE 34th Int. Conf. Data Eng. (ICDE)*, Paris, France, Apr. 2018, pp. 353–364.

[33] R. Li, Y.-P. Lin, Y.-Q. Yi, and Y.-P. Hu, "A privacy and integrity preserving range query protocol in two-tiered sensor networks," *Chin. J. Comput.*, vol. 36, no. 6, pp. 1194–1209, Mar. 2014.

[34] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[35] Z. Chen, P.-H. Ho, and L. Peng, "Optimal hybrid network coding scheme over two-way relaying," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 97–106, Jan. 2019.

[36] Z. Yu, Y. Wei, B. Ramkumar, and Y. Guan, "An efficient signature-based scheme for securing network coding against pollution attacks," in *Proc. IEEE 27th Conf. Comput. Commun. (INFOCOM)*, Phoenix, AZ, USA, Apr. 2008, pp. 1409–1417.

[37] Y. Fan, Y. Jiang, H. Zhu, and X. Shen, "An efficient privacy-preserving scheme against traffic analysis attacks in network coding," in *Proc. IEEE 28th Conf. Comput. Commun. (INFOCOM)*, Rio de Janeiro, Brazil, Apr. 2009, pp. 2213–2221.

[38] T. Elgamal, "A public key cryptosystem and a signature scheme based on discrete logarithms," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 4, pp. 469–472, Jul. 1985.

[39] S. Soro and W. B. Heinzelman, "Cluster head election techniques for coverage preservation in wireless sensor networks," *Ad Hoc Netw.*, vol. 7, no. 5, pp. 955–972, Jul. 2009.

[40] K. A. Hafeez, L. Zhao, J. W. Mark, X. Shen, and Z. Niu, "Distributed multichannel and mobility-aware cluster-based MAC protocol for vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 8, pp. 3886–3902, Oct. 2013.

[41] T. Wang, J. Zeng, Y. Lai, Y. Cai, H. Tian, Y. Chen, and B. Wang, "Data collection from WSNs to the cloud based on mobile fog elements," *Future Gener. Comput. Syst.*, vol. 105, pp. 864–872, Apr. 2020.

[42] T. Wang, Y. Mei, W. Jia, X. Zheng, G. Wang, and M. Xie, "Edge-based differential privacy computing for sensor–cloud systems," *J. Parallel Distrib. Comput.*, vol. 136, pp. 75–85, Feb. 2020.

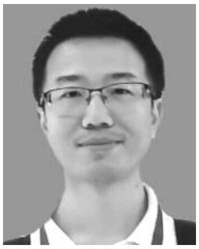
[43] T. Wang, H. Ke, X. Zheng, K. Wang, A. K. Sangaiah, and A. Liu, "Big data cleaning based on mobile edge computing in industrial sensor-cloud," *IEEE Trans. Ind. Informat.*, vol. 16, no. 2, pp. 1321–1329, Feb. 2020.

[44] Y. Wu, H. Huang, Q. Wu, A. Liu, and T. Wang, "A risk defense method based on microscopic state prediction with partial information observations in social networks," *J. Parallel Distrib. Comput.*, vol. 131, pp. 189–199, Sep. 2019.

[45] F. Li. (2018). *Real Datasets for Spatial Databases: Road Networks and Points of Interest*. [Online]. Available: http://www.cs.utah.edu/lifeifei/SpatialDataset.htm

[46] T. Brinkhoff. (2009). *Network-Based Generator of Moving Objects*. [Online]. Available: http://www.fh-oow.de/institute/iapg/personen/brinkhoff/generator/

CHANGLI ZHOU (Member, IEEE) was born in 1985. He received the M.Sc. and Ph.D. degrees from Harbin Engineering University, in 2012 and 2015, respectively. He is currently a Lecturer with the College of Computer Science and Technology, Huaqiao University, Xiamen, China. His recent research interests include cryptography, information security, and privacy preserving.



TIAN WANG (Member, IEEE) was born in 1982. He received the B.Sc. and M.Sc. degrees in computer science from Central South University, in 2004 and 2007, respectively, and the Ph.D. degree from the City University of Hong Kong, in 2011. He is currently a Professor with the College of Computer Science and Technology, Huaqiao University, Xiamen, China. His research interests include wireless sensor networks, fog computing, and mobile computing.

HUI TIAN was born in 1982. He received the Ph.D. degree in computer science from the Huazhong University of Science and Technology, Wuhan, China, in 2010. He is currently a Professor with the College of Computer Science and Technology, Huaqiao University, Xiamen, China. His research interests include network and information security, cloud computing security, and digital forensics.

WENXIAN JIANG was born in 1974. He received the M.Sc. degree from Fuzhou University, Fuzhou, China, in 2006. He is currently an Associate Professor with the College of Computer Science and Technology, Huaqiao University, Xiamen, China. His research interests include network and information security.

ZHIJIAN WANG received the B.Sc. and M.Sc. degrees in computer science and the Ph.D. degree in control theory and control engineering from Central South University, in 1992, 1995, and 2007, respectively. He is currently a Professor with the Information Science School, Guangdong University of Finance and Economics, China. His research interests include system modeling, software engineering, and supply chain management.