南京郵電大學
Nanjing University of Posts and Telecommunications

Lightweight Privacy-Preserving Spatial Keyword Query over Encrypted Cloud Data

汇报人：蒋明峰 指导老师：戴 华

Yang Y, Miao Y, Choo K K R, et al. Lightweight privacy-preserving spatial keyword query over encrypted cloud data[C], 2022 IEEE 42nd International Conference on Distributed Computing Systems (ICDCS 2022), pp. 392-402.

目录

CONTENT

01 背景介绍

02 模型设计

03 实验结果与分析

04 总结与思考

01

背景介绍

背景介绍

➢基于地理位置的服务：

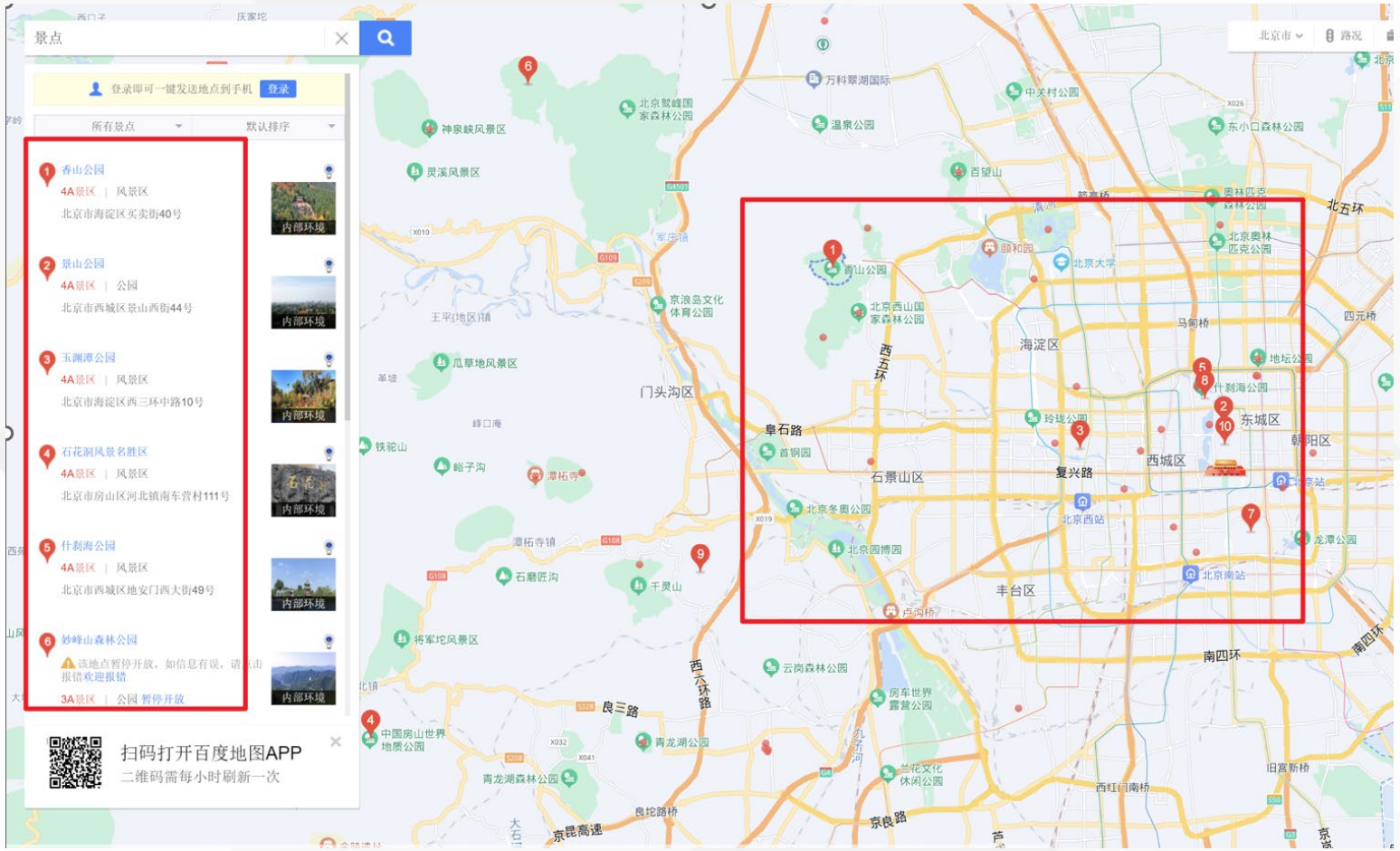
外卖服务

出行服务

社交网络

背景介绍

➢带有地理标签的数据：



背景介绍

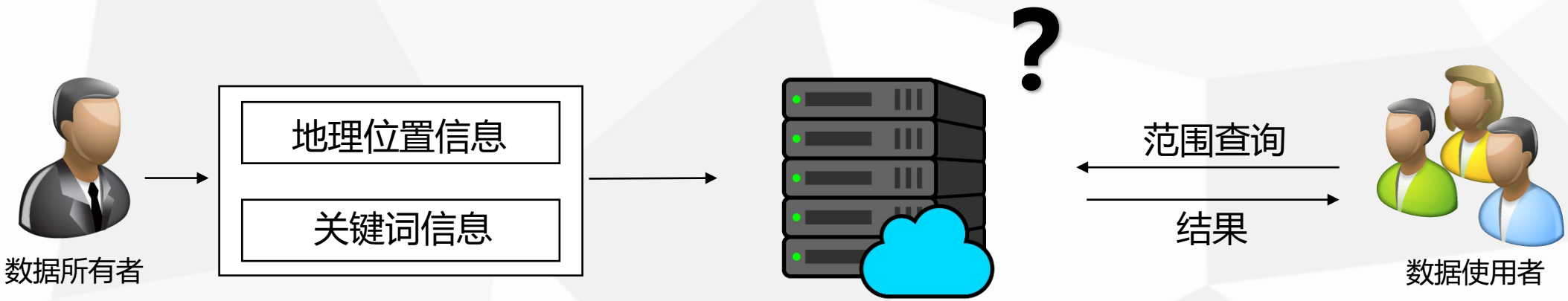
➢查询方式：

范围查询

KNN查询

背景介绍

➢问题描述：



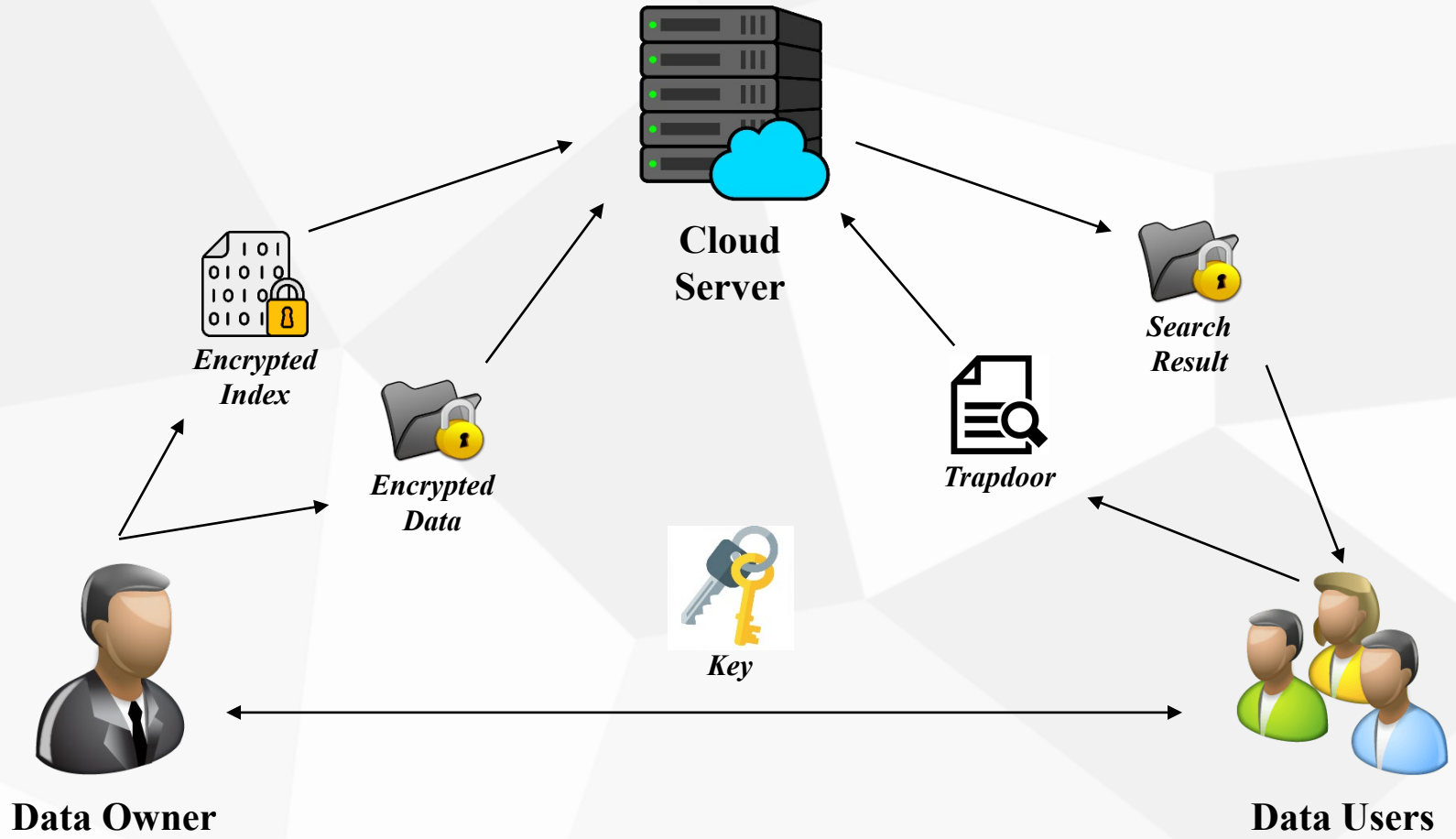
02

模型设计

模型设计

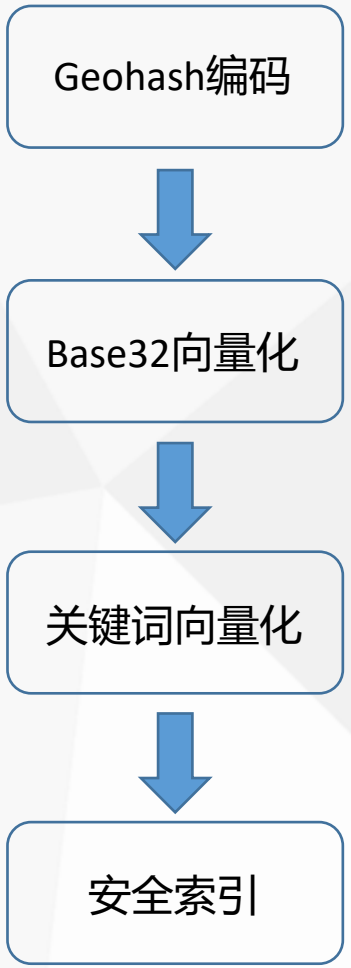
➢范围查询服务模型：

1. 如何将地理位置信息及关键词信息加密并建立安全索引
2. 如何实现高效的范围查询

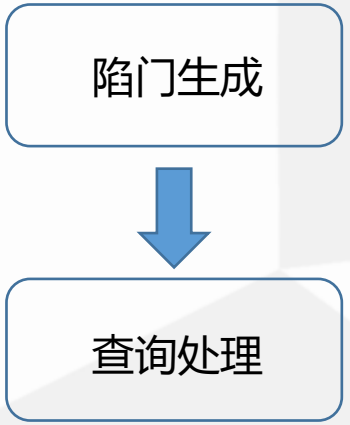


模型设计

数据处理上传阶段



查询处理阶段

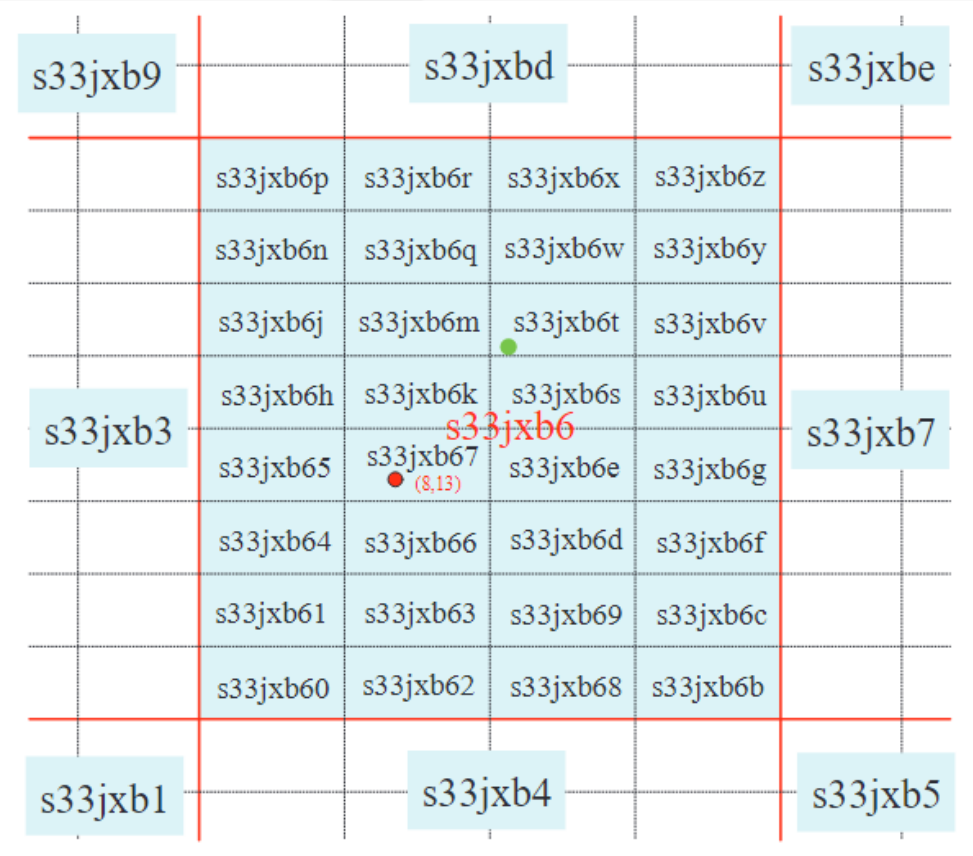


数据处理上传阶段

➢使用Geohash对地理位置信息进行编码：

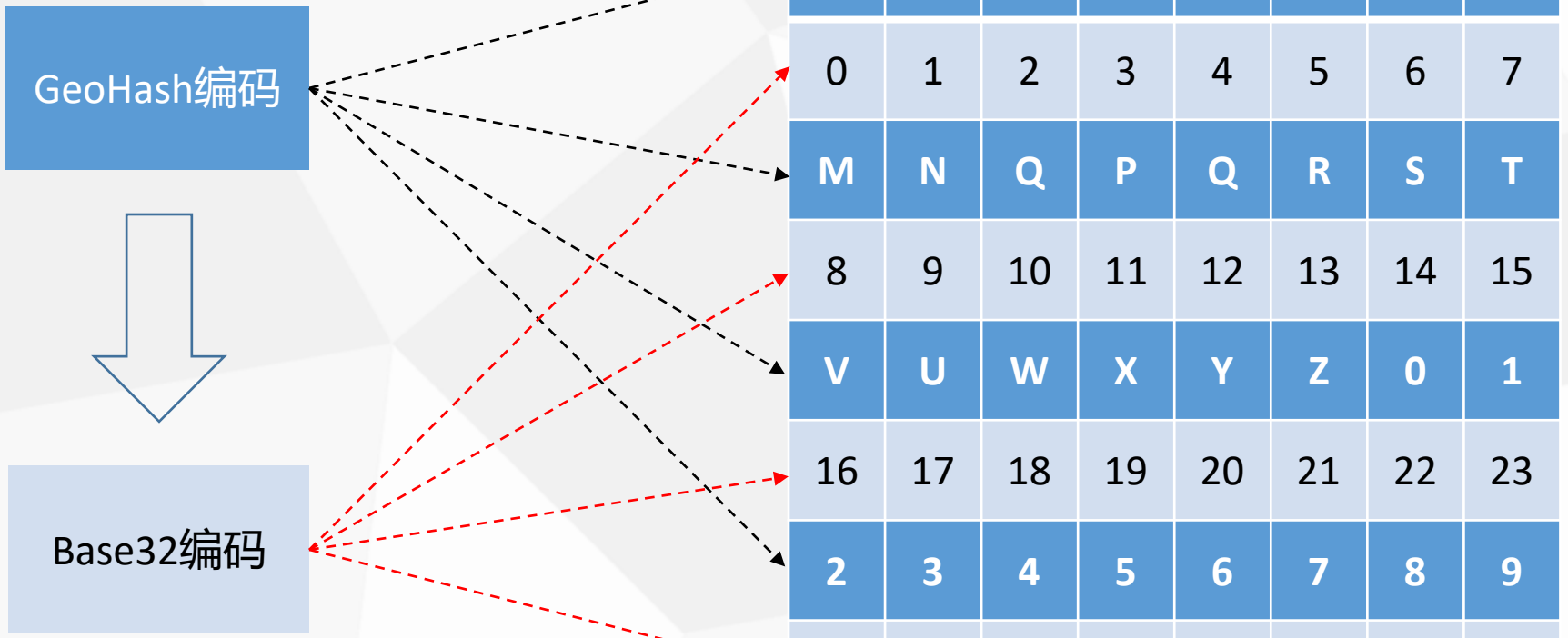
Geohash是空间索引的一种方式，其基本原理是将二维的空间经纬度数据编码为一个字符串，可以把平面递归分解成更小的子块，每个子块在一定经纬度范围内拥有相同的编码。

Geohash编码



Geohash Length	1	2	3	4
Km Error	±2500	±630	±78	±20
Geohash Length	**5**	**6**	**7**	**8**
Km Error	±2.4	±0.61	±0.076	±0.019

Base32编码


GeoHash编码
Base32编码
B C D E F G H J
0 1 2 3 4 5 6 7
M N Q P Q R S T
8 9 10 11 12 13 14 15
V U W X Y Z 0 1
16 17 18 19 20 21 22 23
2 3 4 5 6 7 8 9
24 25 26 27 28 29 30 31

Base32向量化

o.l
(8,13)

Geohash
s33jxb67

(24,3,3,17,29,10,6,7)

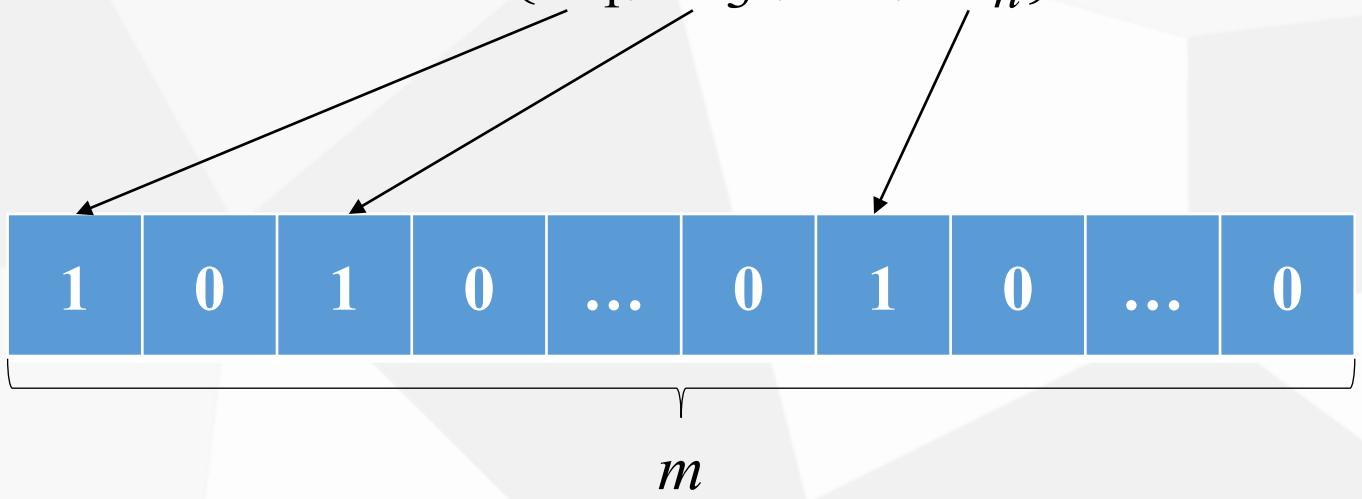
生成8个32维的向量$o.lv_i$

Geohash \ Base32	0	1	2	3	…	6	7	…	23	24	...
s	0	0	0	0	…	0	0	…	0	1	...
3	0	0	0	1	…	0	0	…	0	0	...
…	…										
6	0	0	0	0	…	1	0	…	0	0	…
7	0	0	0	0	…	0	1	…	0	0	…

关键词向量化

$$W = \{w_1, w_2, \ldots, w_m\}$$

$$o.k = \{w_1, w_3, \ldots, w_n\}$$



i.e. $W = \{w_1, w_2, w_3, w_4, w_5\}$，$o.k = \{w_1, w_3\}$，则$o.kv = \{1,0,1,0,0\}$。

安全索引

➢ 存储向量：

存储向量$o.v$为8个$(33+m)$维的向量，$o.v_i = (o.lv_i, o.k, -1)$

i.e. $W = \{w_1, w_2, w_3, w_4, w_5\}$，$o.l = (8, 13)$，$o.k = \{w_1, w_3\}$

o.l
(8,13)

s	0	0	0	0	…	0	0	…	0	1	...	1	0	1	0	0	-1
3	0	0	0	1	…	0	0	…	0	0	...	1	0	1	0	0	-1
…						…						1	0	1	0	0	-1
6	0	0	0	0	…	1	0	…	0	0	…	1	0	1	0	0	-1
7	0	0	0	0	…	0	1	…	0	0	…	1	0	1	0	0	-1

陷门生成

➢ 查询向量：

查询向量$Q.v$为8个(33+m)维的向量，$Q.v_i = (Q.lv_i, Q.k, t)$，t表示$Q.lv_i$和$Q.k$中1的个数

i.e. $W = \{w_1, w_2, w_3, w_4, w_5\}$，$Q.l = (8, 13)$，$Q.k = \{w_2, w_5\}$

Q.l
R = 600m

Geohash
s33jxb

s	0	0	0	0	…	0	0	…	0	1	...	0	1	0	0	1	3
3	0	0	0	1	…	0	0	…	0	0	...	0	1	0	0	1	3
…						…						0	1	0	0	1	3
0	0	0	0	0	…	0	0	…	0	0	…	0	1	0	0	1	2
0	0	0	0	0	…	0	0	…	0	0	…	0	1	0	0	1	2

查询处理

➢ 非对称标量积保持加密（ASPE）：

密钥$sk=\{s, M_1, M_2\}$，向量p, q。

将p加密为$C=(M_1^T p', M_2^T p'')$，其中p利用密钥中的位向量s分成p'和p''。

$$\begin{cases} q'[k] + q''[k] = q[k], & if\ s[k] = 0; \\ q'[k] = q''[k] = q[k], & if\ s[k] = 1. \end{cases}$$

将q加密为$T_Q = (M_1^{-1} q', M_2^{-1} q'')$，其中$q$利用密钥中的位向量s分成$q'$和$q''$。

$$\begin{cases} p'[k] = p''[k] = p[k], & if\ s[k] = 0; \\ p'[k] + p''[k] = p[k], & if\ s[k] = 1. \end{cases}$$

计算结果为

$$C^T \cdot T_Q = \left((p')^T M_1\right) \cdot (M_1^{-1} q') + \left((p'')^T M_2\right) \cdot (M_2^{-1} q'') = p^T \cdot q$$

查询处理

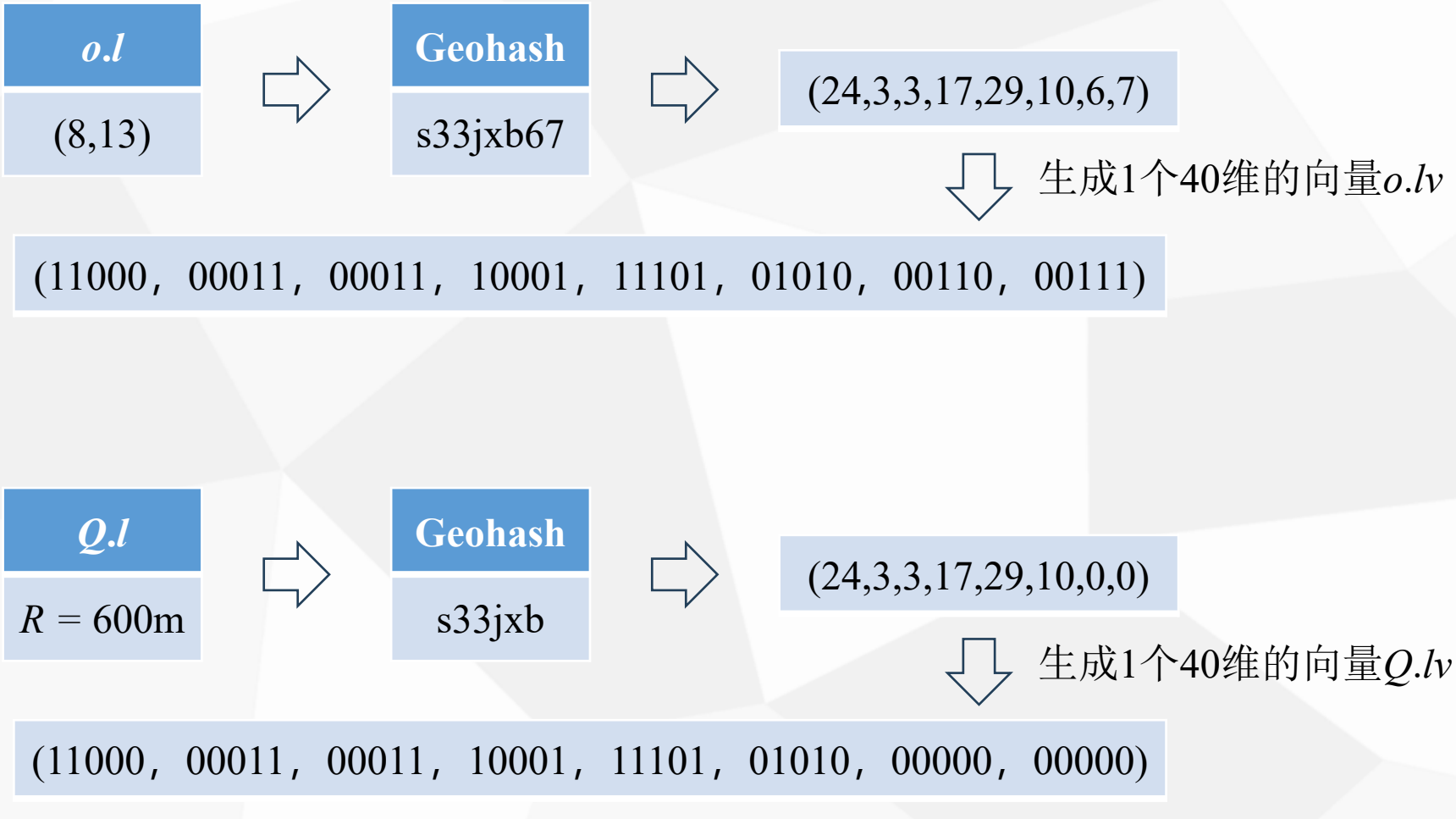
$o.v$

s	0	0	0	0	…	0	0	…	0	1	...	1	0	1	0	0	-1
3	0	0	0	1	…	0	0	…	0	0	...	1	0	1	0	0	-1
…						…						1	0	1	0	0	-1
6	0	0	0	0	…	1	0	…	0	0	…	1	0	1	0	0	-1
7	0	0	0	0	…	0	1	…	0	0	…	1	0	1	0	0	-1

$Q.v$

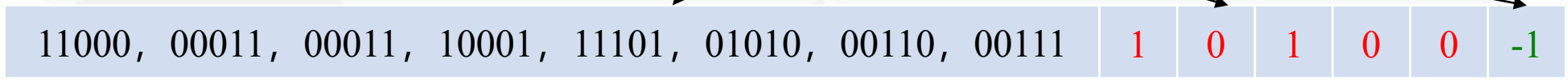
s	0	0	0	0	…	0	0	…	0	1	...	0	1	0	0	1	3
3	0	0	0	1	…	0	0	…	0	0	...	0	1	0	0	1	3
…						…						0	1	0	0	1	3
0	0	0	0	0	…	0	0	…	0	0	…	0	1	0	0	1	2
0	0	0	0	0	…	0	0	…	0	0	…	0	1	0	0	1	2

优化方案

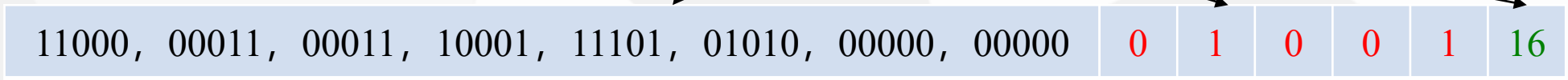

o.l
(8,13)
Geohash
s33jxb67
(24,3,3,17,29,10,6,7)
生成1个40维的向量o.lv
(11000，00011，00011，10001，11101，01010，00110，00111)
Q.l
R = 600m
Geohash
s33jxb
(24,3,3,17,29,10,0,0)
生成1个40维的向量Q.lv
(11000，00011，00011，10001，11101，01010，00000，00000)

优化方案

$o.v = (o.lv, o.k, -1)$



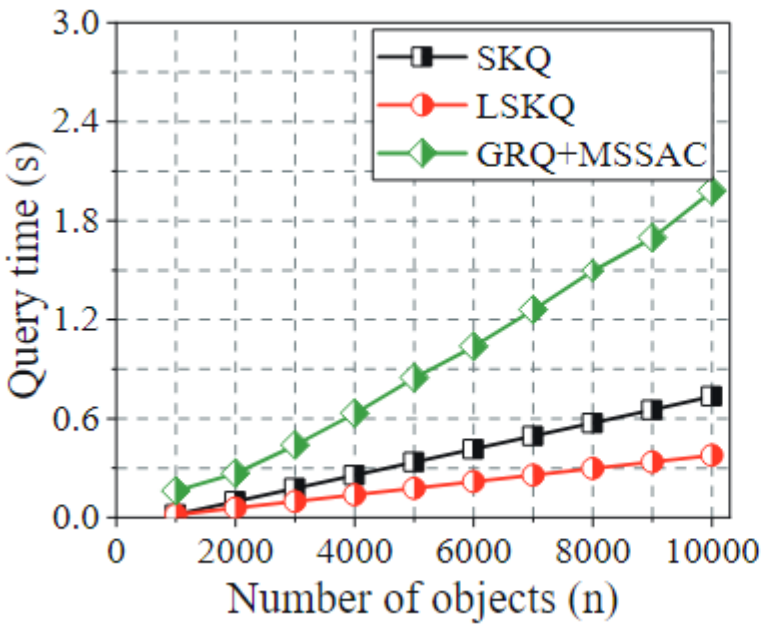
$Q.v = (Q.lv, Q.k, t)$

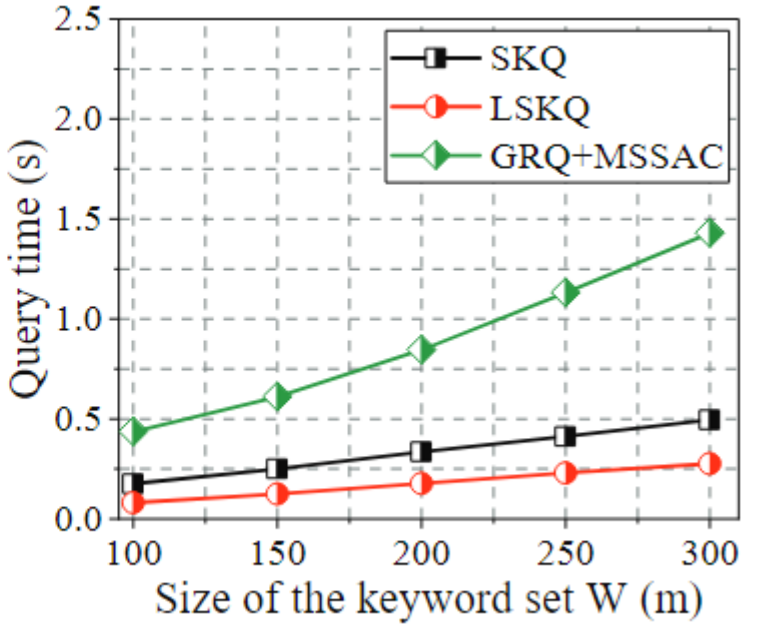


03

实验结果及分析

实验结果


3.0
2.4
1.8
1.2
0.6
0.0
Query time (s)
0
2000
4000
6000
8000
10000
Number of objects (n)
SKQ
LSKQ
GRQ+MSSAC


2.5
2.0
1.5
1.0
0.5
0.0
Query time (s)
100
150
200
250
300
Size of the keyword set W (m)
SKQ
LSKQ
GRQ+MSSAC

04

总结与思考

总结与思考

总结：

1.把地理位置利用Geohash编码并用Base32向量化

2.将关键词向量化后与位置编码拼接

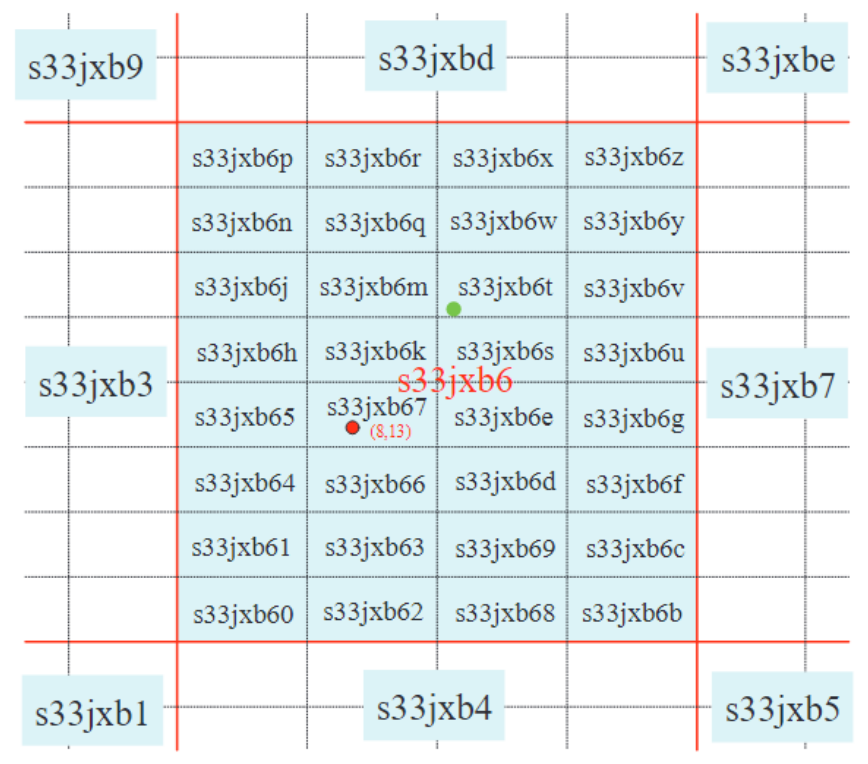
3.利用ASPE实现不可链接性

思考：

1.Geohash的查询范围存在误差

2.优化方案的查询结果有很大的误差

总结与思考



Geohash相邻区域编码差距很大，因此无法跨区域搜索

总结与思考

o_1

77777777

(00111, 00111, 00111, 00111, 00111, 00111, 00111, 00111)

Q

333333

(00011, 00011, 00011, 00011, 00011, 00011, 00000, 00000)

o_2

33333333

(00011, 00011, 00011, 00011, 00011, 00011, 00011, 00011)

$$o_1.v \cdot Q = o_1.v \cdot Q = 12$$

南京郵電大學
Nanjing University of Posts and Telecommunications

敬请各位老师批评指正