International Journal of Grid Distribution Computing
Vol. 8, No. 3, (2015), pp. 351-358
http://dx.doi.org/10.14257/ijgdc.2015.8.3.32

Privacy-Preserving for Check-In Services in MSNS Based on Ant Colony Algorithm

Chen Wen

School of Mathematics and Computer Science, Tongling College, Tongling, P. R. China
tlxychenwen@163.com

Abstract

The sign-in service of Mobile Social Networking Services (MSNS) has a serious issue of privacy disclosure. As to solving this issue, there is a method proposed to protect privacy. This method is to conduct binary coding for sign-in sequence and then construct ant path diagram. The random search, the positive and negative feedback principle of ant path graph are applied to mine frequent item sets. The longest common subsequence is achieved from sign-in sequence and corresponding frequent item sets. Finally, k-anonymous sign-in sequence will be obtained. Based on real data, this method is compared with (k,δ) - anonymity *which proves the accuracy and efficiency of this method through various aspects.*

Keywords: *privacy preservation, mobile social networks, ant colony algorithm, k-anonymous*

1. Introduction

The development of mobile devices and wireless network facilitates the wide application of MSNS. However, sign-in service has a severe risk to disclose privacy. The position of user is a concrete semantic position and it is so specific that the actual position of user is easy to be disclosed. If there are more sign-in positions disclosed, there will be more risks to disclose private positions [1]. In addition, the attacker can analyze and obtain user behavior through sign-in records that the identity of user will be speculated and even the future behavior of user will be forecasted [2].

Privacy protection technology should not only protect user's trajectory privacy, but also provide a higher user experience. The user experience of sign-in service is realized by the effective sign-in position. Therefore, privacy protection algorithm should retain the sign-in position of user so as to increase the successful sign-in rate. At present, there are some research achievements about position privacy protection technology [3-6] and trajectory data privacy protection technology [7-11], but neither of them can be directly applied in MSNS. In recent years, scholars have paid their attention to research on position reasoning attack of MSNS [12], privacy protection of position and absence of MSNS, position privacy protection of nearby service [13-14]. Nowadays, *k*-anonymous privacy has been effectively protected through adopting prefix tree [15] and association graph [16] by researchers in terms of privacy disclosure problem of sign-in service. In order to better improve user experience and protect sign-in privacy, this paper will adopt ant colony algorithm to realize *k*-anonymous privacy protection effect based on the system structure of central server proposed by literature [15]. Ant colony algorithm has a favorable performance to solve complicated optimization problem. The algorithm has been used in the mining of frequent item [18], while there are no research achievements about the application of this algorithm in the field of sign-in privacy protection. The experiment in this paper proves that this algorithm can more efficiently realize *k-anonymous* privacy protection.

ISSN: 2005-4262 IJGDC

2. Generating Algorithm of K-Anonymous Sign-In Sequence

Definition 2.1. Sign-in sequence. The sign-in position of user is recorded as $L_i = (l_i, t_i)$. Meanwhile, l_i represents ID and t_i represents sign-in time. Sign-in record is sequenced according to time which constitutes the sign-in sequence of user.

Definition 2.2. *K*-anonymous sign-in sequence. The privacy protection parameter is *k* and the longest time tolerability isΔt. The sign-in sequence can be called as the *k*-anonymous sign-in sequence of user only if there are $(k-1)$ users have the same sign-in sequence with this user during the time ofΔt.

The maximum possibility for attacker to recognize one particular user from all sign-in sequences will be $1/k$, if sign-in sequence is converted to *k*-anonymous sign-in sequence which can effectively protect the sign-in privacy of users. The generation of algorithm of *k*-anonymous sign-in sequence includes the following procedures:

1. Preprocessing sign-in sequence. In the sign-in sequences of users, the positions which belong to sensitive location privacy region will be deleted;

2. Sequencing the sign-in sequence based on sign-in ID. Conducting binary coding and constructing ant path diagram so as to obtain sign-in *k*-anonymous sequence.

The first procedure is relatively simple that no more details should be set forth. As to the second procedure, *k*-anonymous issue is converted to *k*-frequent item sets mining issue based on ant colony algorithm. The sign-in position whose support degree is lower than *k* will be deleted in order to gain *k*-anonymous sign-in sequence. The concrete procedures of algorithm are shown as following:

2.1. Coding Sign-In Sequence

Constructing bit vector BV_{L_i} for each sign-in record L_i. If sign-in record L_i appears in the sign-in sequence of user u_j, and then the value of bit vector in the position of j is set as 1, otherwise it is set as 0.

2.2. Constructing ANT Path Diagram

If the number "1" of $BV_{L_i} \wedge BV_{L_j}$ $(i < j)$ is not less than the number of privacy protection parameters *k*, then record the values of *i* and *j* and then form a sequence which is denoted as S_2.Each item in the sequence S_2 constitute nodes of ant colony route map, and establish an edge between the two nodes.

Algorithm 1: Constructing ant path diagram

Input: the number of privacy protection parameters *k,* Sign-in sequence L_i.

Output: ant colony route map

```
for(i=1;i<=n;i++){//n is the number of L_i
    for(j=1;j<=n;j++){
        if(Count(L_i, L_j)>=k){//Count(L_i, L_j) used to calculate the number "1" of BV_{L_i} ∧ BV_{L_j}
            establish an edge between L_i and L_j
        }
    }
}
```

2.3. Mining Frequent K Item Sets

Definition 2.3. $p^{\alpha}_{ij}(t)$ denotes the transition probability of ant α from node *i* to node *j* at the time of *t*.

If $j \in allowed_{\alpha}$, $p^{\alpha}_{ij}(t) = \frac{[\tau_j(t)]^{\alpha}[\eta_{ij}(t)]^{\beta}}{\sum_{l \in allowed_{\alpha}} [\tau_l(t)]^{\alpha}[\eta_{il}(t)]^{\beta}}$, otherwise, it equals to 0. (1)

This probability selection formula is controlled by pheromone $\tau_{ij}(t)$ and heuristic value $\eta_{ij}(t)$ together.

Definition 2.4. pheromone $\tau_{ij}(t)$ denotes the degree of expectation for the co-occurrence of *i* and *j* in ant path diagram. After ant moves from node *i* to node *j*, the local pheromone of node *i* is updated to: $\tau_i = (1-\lambda)\tau_i + \lambda\tau_0, \lambda \in (0,1)$ (2)

Definition 2.5. When ant finishes one cycle, the whole pheromones of all nodes are updated to: $\tau_i(t+\Delta t) = (1-\rho)\tau_i(t) + \rho\Delta t_i^{best}(t, t+\Delta t)$ (3).

Meanwhile, if node *i* is included in the optimal path, then $\Delta t_i^{best}(t, t+\Delta t) = \frac{1}{S_{\max}}$, otherwise, it equals to 0.

During the stage of initialization, $\tau_i(0) = \eta_i$. During the process of implementing algorithm, the more ants visit one node, the higher frequency of this node will be. Therefore, pheromone of all sides which access to this node should be increased so as to provide feedback for the following ants.

Definition 2.6. heuristic value η_i is a constant value. in order to take the support degree of node *i* and node *j* of the sign-in sequence into consideration, thus $\eta_{ij} = Sup(i, j)$. As to two nodes which have no borders, $Sup(i, j) = 0$.

The procedures of mining frequent item sets through ant algorithm are shown as following: according to the path search characteristics of ant algorithm and ant path diagram constructed by algorithm 1, nodes of undirected graph will be searched by several ants at the same time and frequent item sets will be generated gradually. In one cycle, ant will start from one node and then walk through all connected roads. Meanwhile, every connected node is only visited for once. A formalized description is shown as follow.

Algorithm 2: Ant algorithm which mines frequent *k* item sets

Initialized parameter: $t = 0$, $\Delta t = 0$, The maximum number of cycles $NC_{\max}$, $\tau_i(t) = \tau_0$, $\rho\Delta t_i^{best}(t, t+\Delta t) = 0$, the number of ants Ant_{num}, the number of privacy protection parameters *k*

Output: frequent item set S_k

```
for(NC=1; NC < NC_max ; NC ++){
    for(α =1; α <= Ant_num ;α ++){// Randomly placing ants on node i
        k = 1;
        S_k(α) = {i};// Frequent item set mined by ant α
        tabu_α = φ;// Emptying the taboo list of ant α
        p = 0;// p is the pointer reference number of taboo list of ant
```

$tabu_{\alpha}(p) = i$;// Letting the pointer of taboo list point to the current position i of ant α

$tabu_{\alpha} = tabu_{\alpha} \cup i$ // Adding the current position i of ant α into taboo list

for($p = 1$; $p <= size(S_k)$; $p++$){//Every ant needs to visit all nodes which are connected to each other

Selecting node j which has the maximum state transition probability according to formula 1

$tabu_{\alpha} = tabu_{\alpha} \cup j$ //Adding position j into taboo list

$allowed_{\alpha} = v - tabu_{\alpha}$ //Updating nodes of feasible paths

if($Sup(S_k(\alpha) \cup j) >= k$){

$S_{k+1}(\alpha) = S_k(\alpha) \cup j$;

$tabu_{\alpha}(p) = j$;//Amending taboo list and letting ant move to new node j

Updating pheromone according to formula 2

$k = k + 1$;

}

}

$t = t + \Delta t$;

}

}

2.4. Solution to K-Anonymous Sign-In Sequence

Definition 2.7. The longest common subsequence. There are two sequences S_i and S_j. If $S_{sub} \subset S_i$ and $S_{sub} \subset S_j$, and there are not subsequence $S'_{sub} \subset S_{sub}$ meeting the above conditions, then S_{sub} is the longest common subsequence of S_i and S_j.

Restoring the sequences in S_n ($n >= k$) to sign-in sequences S'_n. *K*-anonymous sequence will be gained by achieving the longest common subsequence through comparing L_i with S'_n. For instance, frequent item set S_4 is restored to sign-in sequence : S'_4 ={{ l_1, l_2, l_3, l_4},{ l_1, l_3, l_4 ,l_5}}. The sign-in sequence of user is presumed to be $l_1 \rightarrow l_3 \rightarrow l_4 \rightarrow l_5$, and the length of sign-in sequence is presumed to be 4, then the longest common subsequence will be $l_1 \rightarrow l_3 \rightarrow l_4 \rightarrow l_5$ through comparing with S'_4.

Algorithm 3: Obtaining the longest common subsequence

If （n>=k）

{

transform S_n to S'_n

for all L_i do

L'_i=LCS(L_i,$S'_{|Li|}$) // LCS is the longest common subsequence function, $|L_i|$is the length of sign-in sequences, L'_i is the *k*-anonymity sign-in sequence

}

3. Experimental Result and Analysis

The algorithm of this article uses Java, running on Windows XP platform of 4G internal storage with E5800 3.2GHz processor. This experiment applies the real sign-in data on Gowalla [17]. The data set includes the data which are collected for 24 months and the attribute of data set is shown in table 1. *K*-anonymous algorithm (k,δ) - anonymity [9] locates *k*-anonymous set through cluster. Its privacy protection effect is the same to the effect of this algorithm that it can be a comparing algorithm. *K*-anonymous regional sequence of sampling nodes is the output of *k*-anonymous algorithm. Anonymous region cannot be used for sign-in service. This experiment considers the sign-in position covered by anonymous region as an effective output. If anonymous regional sequence covers the sign-in positions of *k* users, then a sign-in sequence can be formed which means that this anonymity is successful.

Table 1. Dataset Properties

check-in number of locations	total number of users	Regional area /km^2	user density
675000	15000	443556	0.03
the average number of check-in	POI average number of check-in	average check-in interval /h	average check-in distance interval /km
44.70	6.20	40.30	19.02

The parameters of this experiment include privacy protection parameter *k*, length parameter of sign-in sequence *len* and the longest tolerance timeΔt. C_S is the success rate of sign in which reflects the proportion of original sign-in position included in *k*-anonymous sign-in sequence. It is measured by formula (4).

$$c_s = \frac{| kChS_iL | - \| kChS_iL | - | ChS_iL \|}{| kChS_iL |} \quad (4)$$

Meanwhile, $| kChS_iL |$ denotes the number of sign-in position in *k*-anonymous sign-in sequence. $| ChS_iL |$ denotes the number of sign-in position in original sign-in sequence.

In this experiment, it will randomly select the sign-in sequences of 4,000 users. The variation of maximum time tolerability is neglected when testing the influence on successful sign-in rate by privacy protection parameter k and length of sequence parameter *len*. The algorithm of this paper is recorded as CA(Check-in privacy perserving based on Ant colony algorithm). The average successful sign-in rate of this algorithm is showed in figure 1.

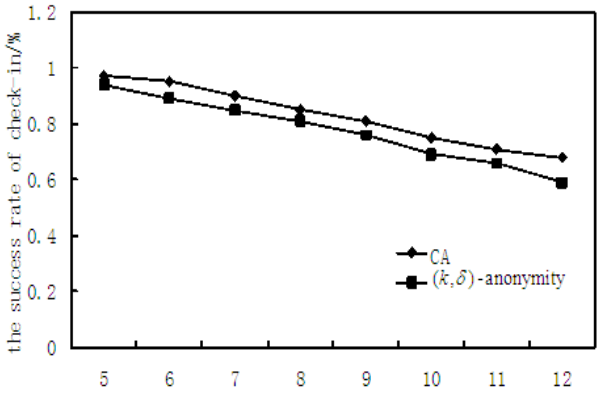
1. Figure 1 (a) shows the influence on successful sign-in rate by the value of privacy protection parameter *k*. When the value of *k* increases from 5 to 12, the successful sign-in rate correspondingly decreases as there are more sign-in positions lost. Meanwhile, the failure rate of (k,δ) - anonymity increases and it leads to the decrease of successful sign-in rate. The successful sign-in rate of algorithm in this paper is relatively high.

2. Figure 1 (b) shows the relation between length of sign-in sequence and successful sign-in rate. *K* is set as 10 and the length of sign-in sequence is from 5 to 30 and the increment is 5 per time. Sign-in sequences are randomly selected to be cut out or complemented. It can be seen that the longer the length of sign-in sequence is, the less identical sign-in sequences will be. This situation will lose more sign-in positions that the successful sign-in rate will be decreased. The increase of *len* has little influence on the successful sign-in rate of (k,δ) - anonymity . It is because that the increase of *len* only has

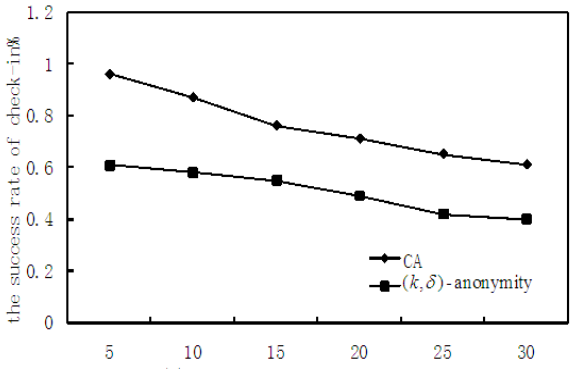
a limited influence on trajectory distance and its influence on cluster result is also relatively small.

3. Figure 2 (c) shows the influence on successful sign-in rate by the longest tolerance time. *K* is set as 10 and the value of the longest tolerabilityΔt is 1 to 5 times of average sign-in time. The longest tolerance time is the cache sign-in time of user. The more the cache sign-in sequences are, the easier to generate *k*-anonymous sign-in sequence will be. When sign-in sequence is relatively less, more sign-in positions will be lost due to the generation of *k*-anonymous sign-in sequence. Therefore, the successful sign-in rate of algorithm has a positive relation with the longest tolerance time. This situation can also be applied to (k,δ) - anonymity .

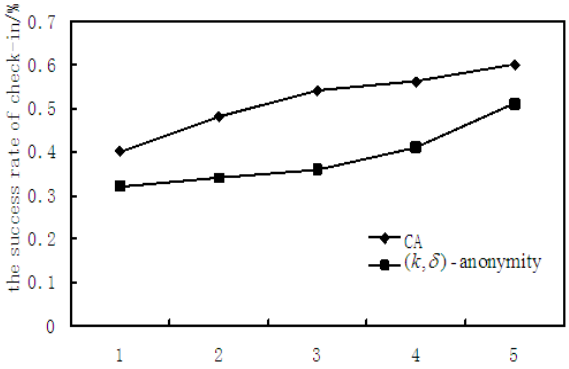
Viewing from the above contrast experiment, the algorithm of this paper is superior to (k,δ) - anonymity at the aspect of successful sign-in rate.



(a) The Parameter K



(b) The Length of Check-In Sequences



(c) The Longest Tolerance Time T

Figure 1. Experimental Results

4. Conclusions

This paper proposes a privacy protection method aiming to the position of fake users and trajectory privacy disclosure issue in MSNS. This method is to convert the *k*-anonymous issue of sign-in sequence into *k*-frequent item sets mining issue based on ant colony algorithm. Sign-in positions whose support degrees are lower than *k* will be deleted and *k*-anonymous sign-in sequence will be achieved. This method simplifies the process of locating trajectory *k*-anonymous set. Furthermore, it not only protects the sign-in privacy of users, but also guarantees a higher successful sign-in rate.

5. Acknowledgement

This work is partially supported by funds from Universities Key Fund of Anhui Province for Young Talents of China under Grant 2013SQRL082ZD and Natural Science Research Universities Key Project of Anhui Province of China under Grant KJ2014A256. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

References

[1] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking", In Proceedings of the 1st international conference on Mobile systems, applications and services, ACM, **(2003)** May, pp. 31-42.
[2] H. Z. M. Xiao-Feng, "A survey of trajectory privacy-preserving techniques", Chinese Journal of Computers, vol. 10, **(2011)**, p. 008.
[3] X. Pan, J. Xu and X. Meng, "Protecting location privacy against location-dependent attacks in mobile services", Knowledge and Data Engineering, IEEE Transactions on, vol. 24, no. 8, **(2012)**, pp. 1506-1519.
[4] M. F. Mokbel, C. Y. Chow and W. G. Aref, "The new Casper: query processing for location services without compromising privacy", Proceedings of the 32nd international conference on Very large data base, VLDB Endowment, **(2006)**, pp. 763-774.
[5] B. Bamba, L. Liu, P. Pesti, *et. al.*, "Supporting anonymous location queries in mobile environments with privacygrid", Proceedings of the 17th international conference on World Wide Web, ACM, **(2008)**, pp. 237-246.
[6] J. Krumm, "A survey of computational location privacy", Personal and Ubiquitous Computing, vol. 13, no. 6, **(2009)**, pp. 391-399.
[7] T. H. You, W. C. Peng and W. C. Lee, "Protecting moving trajectories with dummies", In Mobile Data Management, 2007 International Conference on, IEEE, (**2007)** May, pp. 278-282.
[8] M. Terrovitis and N. Mamoulis, "Privacy preservation in the publication of trajectories", In Mobile Data Management, 2008. MDM'08. 9th International Conference on, IEEE, **(2008)** April, pp. 65-72.
[9] O. Abul, F. Bonchi and M. Nanni, "Never walk alone, Uncertainty for anonymity in moving objects databases", In Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on, IEEE, **(2008)** April, pp. 376-385.
[10] M. E. Nergiz, M. Atzori and Y. Saygin, "Towards trajectory anonymization: a generalization-based approach", In Proceedings of the SIGSPATIAL ACM GIS 2008 International Workshop on Security and Privacy in GIS and LBS, ACM, **(2008)** November, pp. 52-61.
[11] R. Yarovoy, F. Bonchi, L. V. Lakshmanan and W. H. Wang, "Anonymizing moving objects", how to hide a MOB in a crowd, InProceedings of the 12th International Conference on Extending Database Technology, Advances in Database Technology, ACM, **(2009)** March, pp. 72-83.
[12] A. Sadilek, H. Kautz and J. P. Bigham, "Finding your friends and following them to where you are", In Proceedings of the fifth ACM international conference on Web search and data mining, ACM, **(2012)** February, pp. 723-732.
[13] S. Mascetti, D. Freni, C. Bettini, X. S. Wang and S. Jajodia, "Privacy in geo-social networks, proximity notification with untrusted service providers and curious buddies", The VLDB Journal, The International Journal on Very Large Data Bases, vol. 20, no. 4, **(2011),** pp. 541-566.
[14] S. Mascetti, C. Bettini, D. Freni, X. S. Wang and S. Jajodia, "Privacy-aware proximity based services", In Mobile Data Management, Systems, Services and Middleware, 2009, MDM'09, Tenth International

Conference on, IEEE, **(2009)** May, pp. 31-40.
[15] Z. Huo, X. F. Meng and Y. Huang, "Private Check In, Trajectory privacy-preserving for check-in services in MSNS", Jisuanji Xuebao (Chinese Journal of Computers), vol. 36, no. 4, **(2013)**, pp.716-726.
[16] C. Wen, "Privacy-Preserving for Check-In Services in MSNS Based on Association Graph", International Journal of u-and e-Service, Science and Technology, vol. 7, no. 6, **(2014)**, pp. 61-68.
[17] E. Cho, S. A. Myers and J. Leskovec, "Friendship and mobility: user movement in location-based social networks", In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, **(2011)** August, pp. 1082-1090.
[18] T. Li, C. Ling "Ant colony algorithm for mining association rules", Journal of Nanjing University of Posts and Telecommunication (Natural Science), vol. 26, no. 5, **(2006)**, pp. 36-40.

Authors

Chen Wen, is an Associate Professor in the School of Mathematics and Computer Science, Tongling College, Tongling, P. R. China. He holds a master degree in Computer Science and Technology from the Anhui University, Anhui, P. R. China. His previous research areas include privacy preserving.

