A New System for Constructing Location Centric Profiles (LCPS) Of Users That Have Visited Discrete Locations in Geosocial Networks That Guarantees Privacy

M Bharath
M.Tech Student
Department of CSE,
St.Peter's Engineering College,
Hyderabad, TS, INDIA

M Sreenivasulu Reddy
Assistant Professor
Department of CSE,
St.Peter's Engineering College,
Hyderabad, TS, INDIA

N Mahipal Reddy
Assistant Professor
Department of CSE,
St.Peter's Engineering College,
Hyderabad, TS, INDIA

Abstract

Profit is the main participation incentive for social network providers. Its reliance on user profiles, built from a wealth of voluntarily revealed personal information, exposes users to a variety of privacy vulnerabilities. ONLINE social networks have become a significant source of personal information. In this paper, propose to take first steps toward addressing the conflict between profit and privacy in geo social networks. This paper introduce PROFILR, a framework for constructing location centric profiles (LCPs), aggregates built over the profiles of users that have visited discrete locations . PROFILR endows users with strong privacy guarantees and providers with correctness assurances. In addition to a venue centric approach, we propose a decentralized solution for computing real time LCP snapshots over the profiles of co-located users. An Android implementation shows that PROFILR is efficient; the end-to-end overhead is small even under strong privacy and correctness assurances.

Keywords—**Social implications of technology, technology social factors, privacy.**

INTRODUCTION

Geosocial networking (GeoSNs) is a type of social networking in which   geographic services and capabilities such as geo coding and geo tagging are used to enable additional social dynamics. The proliferation of GeoSNs indicates that they're rapidly attracting users. [1] [2] User-submitted location data or geo location techniques can allow social networks to connect and coordinate users with local people or events that match their interests. GeoSNs currently offer different types of servicces, including photo sharing, friend tracking, and "check-ins." However, this ability to reveal users' locations causes newprivacy threats, which in turn call for new privacy-protection methods. The authors study four privacy aspects central to these social networks - location, absence, co-location, and identity privacy - and describe possible means of protecting privacy in these circumstances.Geo location on web-based social network services can be IP-based or use hotspot trilateration. For mobile social networks, texted location information or mobile phone tracking can enable location-based services to enrich social networking.Geo-social networks (GeoSNs) provide context-aware services that help associate location with users and content. Location planning or social-mapping, users are able to search and browse nearby stores, restaurants, etc. Users' venues are assigned profiles and users can rate them, share their opinions and post pictures. These networks use the location of mobile phones to connect users and may also provide directions to and from the venue by linking to a GPS service.[9]

In this paper, we take first steps toward addressing this conflict. Our approach is based on the concept of location centric profiles (LCPs). LCPs are statistics built from the profiles of (i) users that have visited a

certain location or (ii) a set of co-located users.

CONTRIBUTIONS

We introduce PROFILR, a framework that allows the construction of LCPs based on the profiles of present users, while ensuring the privacy and correctness of participants. Informally, we define privacy as the inability of venues and the GSN provider to accurately learn user information, including even anonymized location trace profiles. Verifying the correctness of user data is necessary to compensate for this privacy constraint: users may cheat and bias LCPs anonymously. We consider two user correctness components. First, location correctness, where users should only contribute to LCPs of venues where they are located. This requirement is imposed by the recent surge of fake check ins [5], motivated by their use of financial incentives. Second, LCP correctness, where users should be able to modify LCPs only in a predefined manner.

First, we propose a venue centric PROFILR, that relieves the GSN provider from a costly involvement in venue specific activities. To achieve this, PROFILR stores and builds LCPs at venues.

Furthermore, it relies on Benaloh's homomorphic cryptosystem and zero knowledge proofs to enable oblivious and provable correct LCP computations. We prove that PROFILRsatisfies the introduced correctness and privacy properties.

Second, we propose a completely decentralized PROFILRextension, built around the notion of snapshot LCPs. The distributed PROFILR enables user devices to aggregate the profiles of co-located users, without assistance from a venue device. Snapshot LCPs are not bound to venues, but instead user devices can compute LCPs of neighbors at any location of interest. Communications in both PROFILR implementations are performed over ad hoc wireless connections.

PUBLIC SAFETY & NEWS MEDIA

Most criminal investigations and news events happen in a geographical location. Geo-social investigation tools provide the ability to source social media from multiple networks (such as Twitter, Flickr, and YouTube) without the use of hashtags or keyword searches. Some vendors provide subscription based services to source real-time and historical social media for events.

PRIVACY POLICIES

Some sites, like Facebook, have been scrutinized for allowing users to "tag" their friends via email while checking in ―Check-in vs. Check-out‖An "check-in" is a permission-based network that requires a user to join or sign up. The host is then given permission to access the user's information and to contact him or her. An "check-out" network is defaulted to have the user included in a group. Users must remove themselves from the network if they wish to not be included.

OBJECTIVE

The scope of the project is Toward Preserving Privacy and Functionality in Geosocial Networks using the PROFILER method with two types like venue centric and decentralized.

EXISTING SYSTEMS

Overtly, personal information allows GSN providers to offer a variety of applications, including personalized recommendations and targeted advertising, and venue owners to promote their businesses through spatio-temporal incentives, e.g., rewarding frequent customers through accumulated badges.There exists therefore a conflict.

DISADVANTAGES

1.Providing personal information exposes however users to significant risks, as social networks have been shown to leak and even sell user data to third parties.

2.Without privacy people may be reluctant to use geosocial networks.

3.without user information the provider and venues cannot support applications and have no incentive to participate.

PROPOSED SYSTEMS

This paper introduces PROFILR, a framework that allows the construction of LCPs based on the profiles of present users, while ensuring the privacy and correctness of participants. First, we propose a venue centric PROFILR, that relieves the GSN provider from a costly involvement in venue specific activities. To achieve this, PROFILR stores and builds LCPs at venues. Furthermore, it relies on Benaloh's homomorphic cryptosystem and zero knowledge proofs to enable oblivious and provable correct LCP computations.

Second, this paper propose a completely decentralized PROFILR extension, built around the notion of snapshot LCPs. The distributed PROFILR enables user devices to aggregate the profiles of co-located users, without assistance from a venue device. Snapshot LCPs are not bound to venues, but instead user devices can compute LCPs of neighbors at any location of interest. Communications in both PROFILR implementations are performed over ad hoc wireless connections.

ADVANTAGES

1.Introduce the problem of computing location centric profiles (LCPs) while simultaneously ensuring the privacy and correctness of participants.

2.Propose PROFILR, a framework for computing LCPs. Devise both a venue centric and a decentralized solution. Prove that PROFILR satisfies the proposed privacy and correctness properties.

3.Provide two applications for PROFILR: (i) privacy preserving, personalized public safety recommendations and (ii) privately building real time statistics over the profiles of venue patrons with Yelp accounts.

4.Evaluate PROFILR through an Android implementation. Show that PROFILR is efficient even when deployed on previous generation smart phones.

EXPERIMENTAL BAG ROUND

A core functionality is supported by the most influential geosocial network (GSN) providers, APP [1]and Foursquare [2]. This functionality is simple and general enough to be applicable to most other GSNs (e.g., Facebook Places, Google Latitude). In this model, a provider S hosts the system, along with information about registered venues, and serving a number of users. To use the provider's services, a client application, the ―client‖, needs to be downloaded and installed. Users register and receive initial service credentials, including a unique user id. The provider supports a set of businesses or venues, with an associated geographic location (e.g., restaurants, yoga classes, towing companies, etc). Users are encouraged to report their location, through check-ins at venues where they are present. During a check-in operation, performed upon an explicit user action, the user's device retrieves its GPS coordinates, reports them to the server, who then returns a list of nearby venues. The device displays the venues and the user needs to choose one as her current check-in location.

LOCATION CENTRIC PROFILE

Each user has a profile PU = {pU1, pU2 , .., pUd}, consistingof values on d dimensions (e.g., age, gender, home city, etc).Each dimension has a range, or a set of possible values. Given a set of users U at location L,

the location centric profile at L, denoted by LCP(L) is the set {LCP1, LCP2, .., LCPd }, where LCPi denotes the aggregate statistics over the i –th dimension of profiles of users from U. The intuition behind location privacy (i.e., the first privacy notion given in Section 1) is that users perceive their location as private information. However, they may tolerate that some location information is disclosed if it is sufficiently unlikely that the adversary discovers [4] their precise location. To achieve this result, techniques based on different ideas have been proposed. One idea is to send requests from fake locations together with the request from the real user's location (e.g., [15]). The main problem with the techniques implementing this idea is that a large number of fake request is necessary in order to guarantee privacy protection, while the system costs grow linearly in the number of fake requests. Another solution consists in sending a request (e.g.,a K-NN query) from a fake location and incrementally retrieve results (e.g., NN resources) from the SP until the client can reconstruct the result to the query centered in the real user's location [28]. Privacy is guaranteed because the SP can only discover that the user is located within a region without learning the exact location. The distance between the real user's location and the fake location used in the request determines a trade-o between privacy and performance. Indeed, if the distance is large, the size

5. if ⊥pq∩ C= ∅ or ⊥pq is tangent to C then
6. if |pC| < |qC| then SL := (SL − ϕ) ∪ {[a, b] → p}

else break

7. else
8. let s0, s1 be two points such that ⊥pq∩ C= {s0, s1}
9. if ˆs0 ∈ [a, b] and ˆs1 ∈ [a, b] then

// Assume ˆ s0 < ˆ s1 (the other case is symmetric)
10. if |pCa| < |qCa| then SL := (SL − ϕ)∪

∪{[a, ˆs0] → p, [ˆs0, ˆs1] → q, [ˆs1, b] → p}

// Ca, Cb are the endpoints of arc [a, b]

of region discovered by the SP is also large, but this results in high system costs. These techniques have been applied mostly for LBS performing k-NN spatial queries, and do not apply to proximity detection. A third family of techniques to enforce location privacy is based on the idea of enlarging the user's precise location before it is sent to the SP to a generalized region in order to decrease its sensitivity (among others, [18,25,8]). Some of these techniques are specially designed for proximity services. The main technical problem is how to process spatial queries in which the parameters are generalized regions instead of exact locations. On the other hand, the advantage is that the generalized region can be specifie as a user preference before any information is sent by the client. Indeed, this is the solution we adopt in this paper to protect a user's privacy with respect to her buddies. We actually prove that when a user specifies a generalized region, her buddies do not acquire any location information about that user, except the fact that she is inside the generalized region.

CkNN-Circ(D: the set of objects)
1. for every object p ∈ D do
2. if SL = ∅ then SL := {[0, 2π] → p}

3. else
4. for every interval ϕ ≡ [a, b] → q,ϕ ∈ SL do

11. else SL := (SL − ϕ)∪
∪{[a, ˆs0] → q, [ˆs0, ˆs1] → p, [ˆs1, b] → q}
12. else if ˆs0 ∈ [a, b] or ˆs1 ∈ [a, b] then
// Let only ˆs0 ∈ [a, b] (ˆs1 ∈ [a, b] is symmetric)
13. if |pCa| < |qCa| then SL := (SL − ϕ)∪

∪{[a, ˆs0] → p, [ˆs0, b] → q}

14. else SL := (SL − ϕ) ∪ {[a, ˆs0] → q, [ˆs0, b] →

p}

15. else if |pCa| < |qCa| then
SL := (SL − ϕ) ∪ {[a, b] → p}

16. return SL

CkNN(D: the set of objects)

1. call CkNN-Circ(D)
2. return $\{p : p \in D \land p$ is inside C$\}\cup$
   $\cup\{p : p$ belongs to a mapping of SL$\}$

THE DECENTRALIZATION SERVICE SETTING

A Decentralization service allows its users to publish a resource (e.g., a picture, a text message, a check-in) tagged with the current location and time, as well as a set of users related to the resource. A resource is either tagged automatically (e.g. an integrated GPS can provide location and time), or tagged manually. Since resources and their tags become available to other users as well as to service providers, we are concerned with the privacy violations that the publication can lead to. Formally, a resource r is a tuple:

{Udata; STdata; Contenti}

where the first two elements are meta-data tags with r. U data being a set of identifiers of users, r.STdata being a spatio temporal tag and r.Content being the resource itself. In the following, when referring to a resource r, we also denote with r:Sdata and r:Tdata the spatial and temporal components, respectively, of r:STdata. We assume that all the users in r:Udata are in the location r:Sdata at the time r:Tdata. As an example, recall the user Charlie performing a sta- tus update informing his friends about his presence in the pub together with Alice and Bob. In our formalization, the update is a resource with Alice, Bob, and Charlie as r:Udata, and the location of the pub with the current timeas r:STdata.

We consider techniques for privacy preservation based on the generalization of resources before publication. In partic ular, we consider generalization functions that generalize the spatio-temporal tag of a resource. Formally, STdata for an original resource is a point in the spatio-temporal domain, while STdata for a generalized resource is a 3D volume in the spatio-temporal domain that contains the point of the corresponding original resource.3 In case of generalized resources r0, we denote by r0:Tmax and r0:Tmin the maximum and minimum time instant of r0 Tdata, respectively.

B. Private LCP Requirements

Let k be a security parameter, denoting the level of privacy we need to provide for users at any location. We then define a private LCP solution to be a set of functions, PP(k) = {Setup, Spotter, Check In, PubStats},see Fig.1 Setup is run by each venue
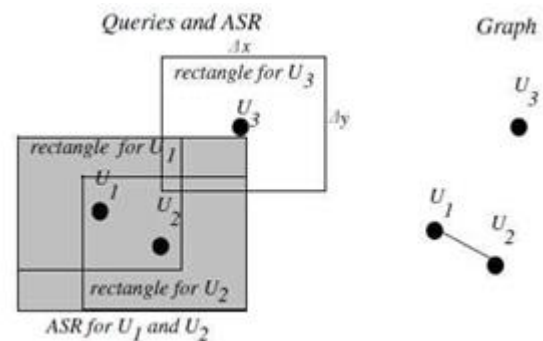
whereuser statistics are collected, to generate parameters for user check-ins. To perform a checkin, a user first runs Spotter, to prove her physical presence at the venue. Spotter returns error if the verification fails, success otherwise. If Spotter is successful, Check In is run between the user and the venue, and allows the collection of profile information from the user.

Specifically, if the user's profile value v on dimension D falls within the range Ri , the counter c_i is incremented by 1. Finally, PubStats publishes collected LCPs. In the following, we use the notation Prot (P1(args1), .., Pn(argsn)) to denote protocol Prot run between participants P1, .., Pn, each with its own arguments. Let CV be the set of counters defined at a venue V. We use .CV to denote the set of sets derived from CV as follows. Each set in .C V differs from CV in exactly one counter, whose value increments the value of the corresponding counter in CV . For instance, if CV = {2, 5, 9}, then .C V ={{3, 5, 9}, {2,

ALGORITHM: ANONYMIZATION ALGORITHM

INPUT: T1, T2 a k-privacy requourement, a taxonomy tree for each categorical attribute in xn.

OUTPUT: a generalized T2 satisfying the privacy require ment.

1. Generalize entry value of Ai to Anywhere $Ai \in Xi$

2. While there is a valid candidate in ^{UU}ut, do

3. Find the paire of near root (xi)from Úcut.

4.   Specialized or on t2 and remove Xifrom Úcut.

5. Replace new (xi) and the valid status of xi for all in Úcut.

6. Out put the generalized T2 and Úcut.

User Interface

(b) Example of NN query
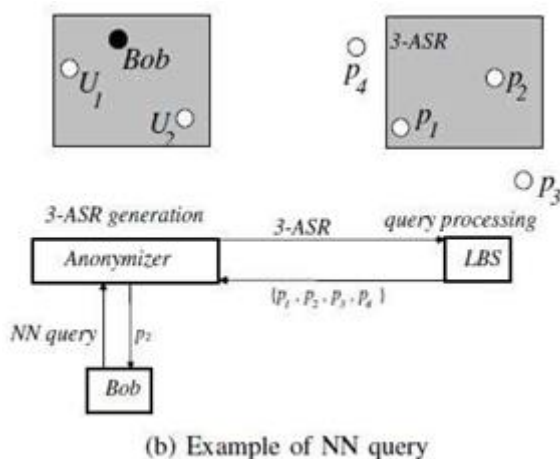
User use the applications for the incentive and some application implementation.In this process user search the application from the database.In that time user location stored into the database. Provide two applications for PROFILR: (i) privacy preserving, personalized public safety recommendations and (ii) privately building real time statistics over the profiles

of venue patrons with Yelp accounts. Evaluate PROFILR through an Android implementation. Show that PROFILR is efficient even when deployed on previous generation smartphones.

Users can verify the results of their queries, relying only on their trust of the data owner. In addition to assuming a different environment, PROFILR does not assume venue owners to be trustworthy. Users have a profile that allows the private matching of relevant ads. While PROFILR can be used to privately provide location centric targeted ads, its main goal is different - to compute location (venue) centric profiles that preserve the privacy of contributing users. By Method--Thompson et. al. approach, zero-knowledge process.
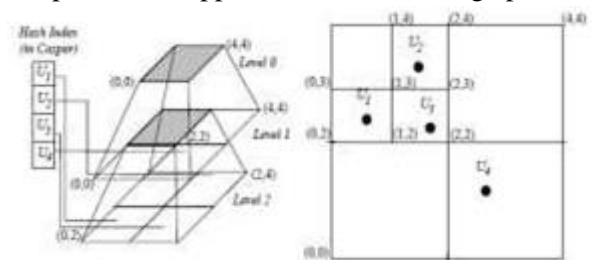

CHECK-IN PROCESS

When user use the application,In that time server check the lcp for the user correctness.This process perform with some actors like Setup, Spotter, Check In, PubStats.This paper use one of the protocols proposed in to verify the location claims of users checking-in. This method assume a honest challenger, who does not run Spotter and Check In twice for the same (user, epoch) pair. Otherwise, the use of the signed pseudonyms provides an advantage to some process. Note that if pseudonyms are not used, this requirement is not necessary. No identifying information is sent by users during the Spotter and Check In procedures. Users are encouraged to report their location, through check-ins at venues where they are present. During a check-in operation, performed upon an explicit user action, the user's device retrieves its GPS coordinates, reports them to the server, who then returns a list of nearby venues. The device displays the venues and the user needs to choose one

as her current check-in location by Method—SPOTRV protocol, threshold secret sharing (TSS), Spotter protocol

EXPERIMENTAL EVALUATION

This section evaluates the proposed anonymization and query processing algorithms. We implemented prototypes for both the anonymizer and the LCP using JAVA. All experiments were executed on an Intel Xeon 2.8GHz machine with 2.5GB of RAM and Linux OS/windows 7. Our workload for user positions and landmarks/points of interest consists of the NA dataset [30], which contains 569K locations on the continent (Figure 16). Performance is measured in terms of CPU time, I/O time and communication cost. At the anonymizer we employed mainmemory structures, therefore we measured only the CPU time. At the LCP, we used an R*-Tree and measured the total time (i.e., I/O and CPU time); in all experiments we maintained a cache with size equal to 10% of the corresponding R*-Tree. The communication cost was measured in terms of number of candidates sent from the LCP back to the anonymizer.
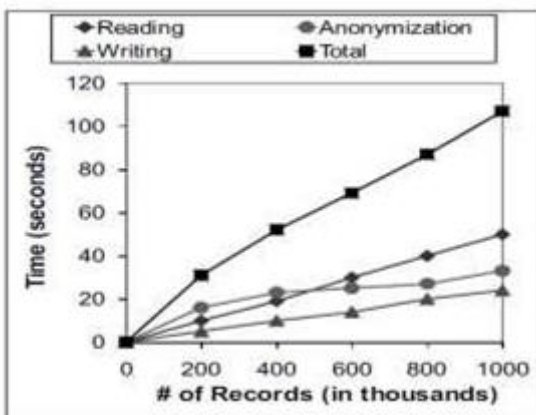
We propose to use PROFILR to build finer grained personalized safety recommendations, with privacy. PROFILR divides the safety index interval ([0, 1]) into sub-intervals, and associates a counter with each. PROFILR enables then a set of users t privately and correctly compute the distribution of their safety index values.
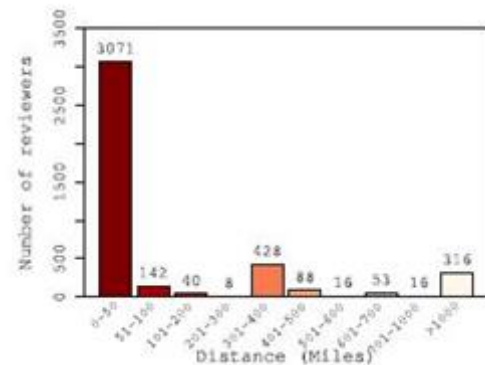
The computation overhead of Check In is $TCI = bTRE + TZK$, where TRE is the Benaloh re-encryption cost and TZK is the overhead of the ZK-CTR protocol. The formula does not consider the cost of modular multiplication, random number generation and random permutation operations, that are neglibile compared to the other costs. Given s, the number of rounds of ZK-CTR, $TZK = 2sbTRE + sbTRE + s2\ bTRE = 72\ sbTRE$. The communication overhead is $Tcom\_CI = bN + Tcom\_ZK$. The communication cost of ZK-CTR, $Tcom\_ZK$ is $s(2bN + 12\ 4bN + 122bN) = 5sbN$.

CONCLUSION:

In this paper we have proposed PROFILR, a framework and mechanisms for privately and correctly building location centric profiles. We have proved the ability of our solutions to satisfy the privacy and correctness requirements. We have introduced two applications for PROFILR. We have shown that PROFILR is efficient, even when executed on resource constrained mobile devices.

FUTURE ENHANCEMENTS

There are some potential future directions of this work. In particular, besides the power-law distribution, it is promising to consider other methods for modeling the geographical mobility patterns of users. Moreover, it is also interesting to explore the performance of different combinations of geographic influence and social influence in addition to their product. An interesting direction for future work is to process Geo-Social queries based on the trajectories of the mobile users.

The main challenge is how to calculate the geo-distance between users based on the history of the locations, not only the current locations.

REFERENCES:

1. M. Gruteser and D. Grunwald, ―Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking.‖ in Proc. of USENIX MobiSys, 2003, pp. 31–42.

2. B. Hoh and M. Gruteser, ―Protecting Location Privacy Through Path Confusion,‖ in Proc. of SecureComm, 2005, pp. 194–205.

3. H. Hu and D. L. Lee, ―Range Nearest-Neighbor Query,‖ IEEE TKDE,vol. 18, no. 1, pp. 78–91, 2006.

4. P. Kamat, Y. Zhang, W. Trappe, and C. Ozturk, ―Enhancing Source-Location Privacy in Sensor Network Routing,‖ in Proc. of ICDCS, 2005,pp. 599–608.

5. K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, ―Incognito: EfficientFull-Domain K-Anonymity.‖ in Proc. of SIGMOD, 2005, pp. 49–60.

6. A. Machanavajjhala, J. Gehrke, D. Kifer, and M.Venkitasubramaniam,―l-Diversity: Privacy Beyond k-Anonymity.‖ in Proc. of ICDE, 2006.

7. A. Meyerson and R. Williams, ―On the Complexity of Optimal Kanonymity,‖in Proc. of

ACM PODS, 2004, pp. 223–228.

8. M. F. Mokbel,W. G. Aref, and I. Kamel, ―Analysis of Multi-DimensionalSpace-Filling Curves,‖ GeoInformatica, vol. 7, no. 3, pp. 179–209, 2003.

9. M. F. Mokbel, C. Y. Chow, and W. G. Aref, ―The New Casper: Query Processing for Location Services without Compromising Privacy,‖ in Proc. of VLDB, 2006, pp. 763– 774.

10. B. Moon, H. Jagadish, and C. Faloutsos, ―Analysis of the Clustering Properties of the Hilbert Space-Filling Curve,‖ IEEE TKDE, vol. 13, no. 1, pp. 124–141, 2001.

11. P. Samarati, ―Protecting Respondents' Identities in Microdata Release.‖IEEE TKDE, vol. 13, no. 6, pp. 1010– 1027, 2001.

12. H. Samet, The Design and Analysis of Spatial Data Structures.Addison-Wesley, 1990.

13. L. Sweeney, ―k-Anonymity: A Model for Protecting Privacy,‖ Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 10, no. 5,pp. 557–570, 2002.

14. Y. Tao and D. Papadias, ―Historical spatio-temporal aggregation,‖ ACMTrans. Inf. Syst., vol. 23, no. 1, pp. 61–102, 2005.