A location privacy preserving algorithm based on linkage protection

CONFERENCE PAPER · JANUARY 2011

DOI: 10.1109/ICITIS.2010.5688758 · Source: IEEE Xplore

Z.P. Jin
Deggendorf University of Applied Sciences
3 PUBLICATIONS **7** CITATIONS
SEE PROFILE

Jian Xu
Hangzhou Dianzi University
18 PUBLICATIONS **34** CITATIONS
SEE PROFILE

Ming Xu
Hangzhou Dianzi University
42 PUBLICATIONS **89** CITATIONS
SEE PROFILE

Ning Zheng
Hangzhou Dianzi University
40 PUBLICATIONS **81** CITATIONS
SEE PROFILE

A Location Privacy Preserving Algorithm Based on Linkage Protection

Z.P. Jin, Jian Xu, Ming Xu, Ning Zheng
Computer Science Department of
Hangzhou Dianzi Technological University
HangZhou, China, 310085
jin_z_p@hotmail.com, {jian.xu, mxu, nzheng}@hdu.edu.cn

***Abstract*—The key of location privacy preserving is to protect the unlinkability between location and identity. Most of existing algorithms focus on *location protection* and *identity protection* separately. That will leads to decreased service quality, authentication, and auditability. In this paper, we propose a new algorithm which uses a variable-length anonymous communication path to protect the linkage between location and identity when users publish their location information. And give out detailed introduction and analysis of the algorithm. We evaluate the performance of the algorithm via simulation and show that it significantly increase anonymity, scalability and auditability at last.**

Keywords-location privacy preserving; VANET;

I. INTRODUCTION

Applications in Vehicular Ad hoc Network (VANET) have greatly enriched people's driving life. Just stay in the car, people can query real time traffic information, browse the Internet and pay tolls. Some of these applications ask people to reveal precise location which means exposing their privacy. Privacy is a fundamental right of people which has been codified in law. In the VANET, lots of algorithms have been proposed to protect the location privacy, and the key of location privacy preserving is to protect the unlinkability[1] [2] [3] between location and identity. The literature [4] surveys the algorithms to two types: location protection and identity protection. Location protection algorithm puts attention to the location. It uses obscure location to instead of precise location. It achieves unlinkability but reduces service quality. Identity protection algorithm has a different focus on identity. It tries to keep the real identity distinguishable from others but brings difficulty for authenticating (need real identity).

In this article, we propose a linkage protection algorithm to achieve the unlinkability requirement. The algorithm reveals identity and location but protect the linkage relationship between them. It can achieve good service quality with precise location and good authentication with real identity. The linkage is broken by a variable-length anonymous communication path before the message published. The algorithm can offer a better internal anonymity (anonymity to internal collaborator) and external anonymity (anonymity to external observer). The rest of the paper is organized as follows: Section 2 surveys the related work. In section 3, we discuss some essential issues about location privacy preserving. We expand our algorithm in section 4. Section 5 gives out analysis and experimental results. Finally, we conclude the paper in section 6.

II. RELATED WORK

For achieving the unlinkability between location and identity, there are two types of algorithms: location protection and identity protection.

Location protection algorithm focuses on the location. Kido H et al propose a dummy strategy [5]. It is a location protection algorithm. It offers redundant location when user accesses the LBS server. But it reduces the usability of the data collected from users. Obfuscation [6] has a similar strategy and it offers a regional cloaking location instead of the precise location. The k-anonymity strategy [7] is also a location protection strategy. It uses a regional location which cloaks all of the k nodes instead of precise location. All of those strategies have good authentication but lower quality of service because the server unable to collect precise location

Identity preserving means keeping the identity from being eavesdropped. Anonymity is the most common strategy but weak for authentication. Gruteser M and Grunwald D [8] firstly propose the k-anonimity model through spatial and temporal cloaking algorithm. It composes a k nodes group and keeps the initiator indistinguishable from other k-1 nodes. ChiYin Chow et al improved the model in Peer-to-Peer architecture [7]. Pseudonymity [9] is a special anonymity strategy using a pseudonym instead of user's identity. It gives consideration to anonymity and authentication but is still not reliable under data mining technology. Beresford and Stajano [10] give another model——Mix-zone. It defines two zones: application zone and mix zone. Users keep silent and anonymous to LBS server in mix zone. The interaction to server only occurs in the application zone. The Mix-zone is effective for preserving identity information but it is hard to be deployed in large scale system.

III. PRIVACY ISSUES IN VANET

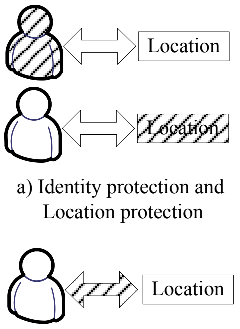
A. Privacy Threats

There are many practical applications in VANET which have great improved traffic efficiency and safety. Those applications can be divided into two types: *comfort-related application* and *safety-related application*. Comfort-related applications are most used to improve the traveling comfort or traffic efficiency, such as distributed traffic information system, VOD, internet access, online games and so on. Safety-related applications is related to life-critical situation and has a higher security requirement, such as distance warning and collusion avoidance are two common safety-related applications.

However, all of the applications must face to a problem that protects the location privacy of a vehicle. Most of the VANET applications may ask for precise location data to promote the service quality. But a vehicle is private and its location is privacy related information. The correlation of someone's location and other personal data may expose much useful information to attacker. For examples, a vehicle appears frequently in a hospital can get such an inference that the driver may be ill; if a vehicle just parked before a bank several minutes ago, the driver may have some crash. Some of the information may threat people's security.

B. Privacy Protection Policies

The location data processed by computer in VANET application can be denoted by a tuple $[id, loc, other]$ with identity, location and ohter accessorial information. For an example, a tuple [Bob, hospital, other] means Bob drives around the hospital. If we publish the linkage of Bob and hospital, the privacy will be exposed. The purpose of location privacy preserving algorithm, indeedly, is to protect the linkability between id and loc.



a) Identity protection and Location protection

b) Link protection

Figure 1. Different location privacy preserving policies

Fig 1(a) shows a simple diagram for two policies. They have a different focuses on identity and location (denoted by bias area) to achieve the unlinkability requirement. However, both of them have some defects. Keeping identity anonymity (identity protection) brings authentication difficulty and will be hard to pay for some charged service. Obfuscated location data (location protection) reduces the service quality dramatically. So we want a new strategy with good authentication and privacy protection.

In this paper, we propose a new strategy— linkage protection—to achieve the unlinkability requirement. Our new strategy pays attention to protect the linkage relationship between location and identity (Figure 1(b)). We reveal real identity and location data in every tuple, but broke the one-to-one relationship between them via an anonymous path.

C. P2P Spatial Cloaking Algorithm

P2P spatial cloaking algorithm is proposed by ChiYin Chow et al. It is a typical algorithm protecting both location and identity. The algorithm covers a VANET to a peer-to-peer network. The main idea of the algorithm is that an originator will find an agent peer to reveal its message. To achieve the privacy requirement, the agent will discover k peers to constitute a group. So the originator can't be distinguished from other k-1 peers (focus on identity). The group will cover an area which should be more than a minimum area A_{min} (focus on location). It can still keep anonymity when the originator in a high peer density area. Figure 2 shows a running example of P2P spatial cloaking algorithm. A is the originator and finds C as the agent. Then C discovers A and B to constitute a group (k=3) and covers a group area A ($A > A_{min}$). C will be a center peer in the group and reveal treated messages for both A and B.

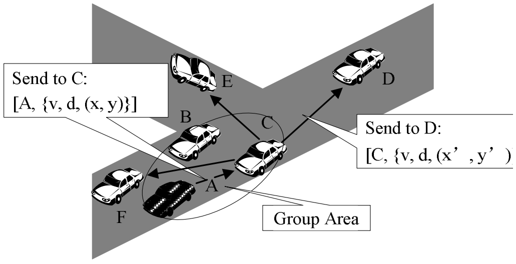


Figure 2. A running example of P2P spatial cloaking algorithm

Originator A sends messages $[A, \{v, d, (x, y)\}]$ to C. A denotes the identity while $\{v, d, (x, y)\}$ denotes a message with velocity, direction and location (x, y). Then C uses a treated location (x', y') (the group area location) to replace original location and reveals treated message $[A, \{v, d, (x', y')\}]$ to others for A. So the group has an anonymity parameter with k and A to the outside of the group.

However, the P2P spatial cloaking algorithm still has some defects. Firstly, the P2P algorithm replaces the exact peer location with a group area location. It reduces the data availability collected by server and influences the service quality dramatically. Another defect is about the agent. It brings a big risk on the center peer. Once the peer is taken up by a collaborator, all of the messages will be dangerous. And it also breaks the equality about the right and duty of every peer. The center peer is in charge of more work than others.

IV. ALGORITHM

A. Linkage Protection Algorithm

VANET is a distributed, self-organize and multi-hops network. The vehicles communicate with each other via a short distance wireless device. But the communication is restricted by the transmission range. All of the vehicles are divided into several groups because of the transmission range restriction (Figure 3).

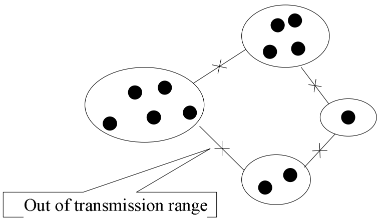
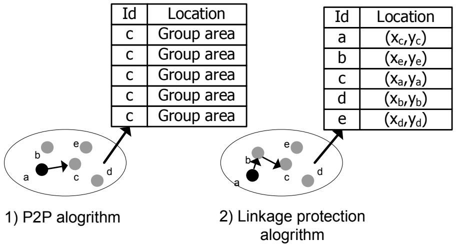


Figure 3. The division of groups

In some safety-related applications (e.g. distance warning), every node is required to reveal the real-time movement status, including velocity, direction, location, acceleration and so on. Revealing real-time and exact location means exposing privacy. In this paper, we propose a linkage protection algorithm. The linkage protection algorithm finds an anonymous path to publish message while the P2P algorithm uses an agent. You can get a macroscopic view in Figure 4. The P2P algorithm publishes group area with the agent's id with a losing of data usability. But the linkage protection algorithm publishes a confused relationship table of real id and real location. The peers in the group are organized by several anonymous paths.



Id	Location
c	Group area
c	Group area
c	Group area
c	Group area
c	Group area

1) P2P alogrithm

Id	Location
a	(x_c, y_c)
b	(x_e, y_e)
c	(x_a, y_a)
d	(x_b, y_b)
e	(x_d, y_d)

2) Linkage protection alogrithm

Figure 4. A macroscopic view of P2P and linkage protection algorithm

The linkage protection algorithm can simply be divided into three phases: authentication phase, anonym phase and publish phase.

Authentication phase. Every vehicle will be authenticated when connects to the self-organized VANET first time. We should ensure every identity real, because several applications are life-critical and all of the vehicles should be charged for their published messages.

Anonym phase. In this phase, we create a variable-length anonymous communication path with several forward peers. The communication among those peer is secret via encryption techniques. We protect sender anonymity [11] against the collaborator attacker among the forward peers. Figure 5 shows a running example for the algorithm. Original sender set a random anonymous path's length as 2 firstly and sends a tuple [A, {v, d, (x, y)}, 1] with identity A, message {v, d, (x, y)} and rest length of anonymous path 1 (subtracted itself by 1) to next forward peer B. Then B receives information from A but it doesn't know the message is about A (because B can't distinguish that A is the original sender). B will do its duty that subtract rest-length with 1 and forward the message to next peer C. The anonymous won't completely created until a peer receives a tuple with 0 rest-length.

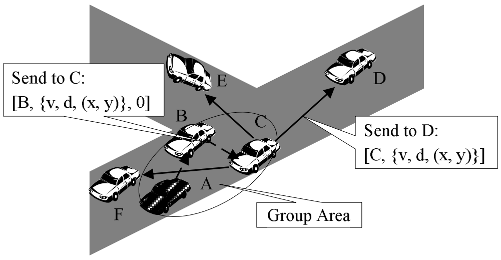
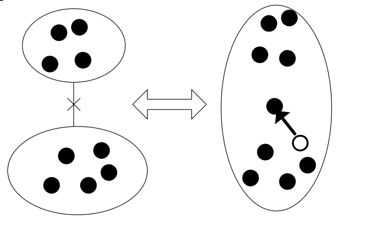


Figure 5. A running example of linkage protection algorithm

Publish phase. After the anonymous path is created, the last forward peer will publish the information to other peers with single hop or multi-hops public broadcasting. Then every peer gets the message {v, d, (x, y)} but can't distinguish whom the message is from, because the publisher is not the original sender. The linkability between the message and the original sender is protected.

B. Peer Selecting Strategy

Because of the fast moving of vehicles, the group topology changes dramatically. Two separate groups may be merged when a moving peer connects both of them while a group can divided into two separate groups when move out of the transmission range (Figure 6). The merge and division take place frequently. And we need to maintain the confused relationship table quickly. The table is created by the constructing of all anonymous paths. So we need an effective peer selecting strategy to maintain the table when creates the anonymous path.



a) separate groups b) merged group

Figure 6. The merge of two separate groups

Common random peer selecting policy has three types policies (Figure 7). All of them have different group area (bias area) growth rate. Without a doubt, a) in Figure 6 has the maximum group area growth rate, next is c and b is the minimum. Because the third peer of a) is the most far away from first and second peer. A new peer which is more far away from the group area centroid will have a bigger growth rate. The location of the peer should be

$$(x_i, y_i) \leftarrow Max\{d_1, d_2, \cdots d_i \cdots\}$$

with $d_i = \sqrt{(x_i - X)^2 + (y_i - Y)^2}$, (X,Y) denotes the centroid location while (x_i, y_i) is the location of every vehicle. After the new peer adds to the group, the group area centroid location will be updated as follow:

$$(X', Y') = \frac{n}{n+1}(X, Y) + \frac{1}{n+1}(x_i, y_i)$$

with *n* is the path length. So we can select a next peer with big area growth rate.

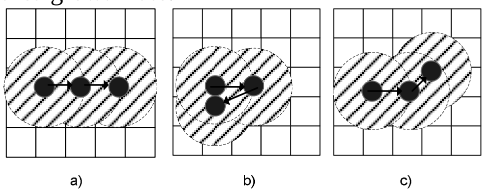


Figure 7. Three types of general peer selecting policy.

V. Analysis and experimental evaluation

We will give our analysis and experimental evaluation in anonymity, scalability and auditability, respectively.

A. *Internal Anonymity*

Existing literatures [12] propose a group concept but they don't analysis the internal anonymity. They presume that all peers in the group are trusty. But sometimes there are some collaborators in the group. The sensitive message will be eavesdropped. Our paper is the first time to analysis the anonymity to internal collaborator.

We use the probability of that attacker successfully figure out the originator to evaluate the anonymity. We denote H as a event that the attacker figure out the originator and P(H) means the probability. We assume that there are c collaborators amount n (n>c) mobile peers. We discuss the internal anonymity about P2P algorithm and linkage protection algorithm.

In P2P algorithm the information is forwarded by an agent. So the attacker can know who is originator only if the agent is a collaborator. The probability can be calculated as:

$$P_1(H) = \frac{c}{n}$$

In linkage protection algorithm, the information is disseminated by a variable-length anonymous path. If an attacker is in the anonymous path, it will have a probability to infer that the *i*th node from itself is the originator. The probability can get from an exponential distribution $\varphi(i) = \lambda e^{-\lambda i}$ $(\lambda > 0)$. The probability of attacker take the *i*th position can be calculated as $\rho(i) = (\frac{n-c}{n})^{i-1}\frac{c}{n}$. The probability of H_i (the event that the *i*th node successfully figure out the originator) is: $P(H_i) = \varphi(i)\rho(i)$ and the probability of originator is figured out by attacker can be calculated as follow:

$$P_2(H) = \sum_{i=1}^{\infty} \lambda e^{-\lambda i}(\frac{n-c}{n})^{i-1}\frac{c}{n} = \frac{c\lambda e^{-\lambda}}{n-(n-c)e^{-\lambda}}$$

We can proof mathematically that $P_2(H) < P_1(H)$ when $(\lambda > 0)$. So the linkage protection algorithm has a lower probability of that originator will be figured out by attacker, which means it has better internal anonymity.

B. *External Anonymity*

The external anonymity is used to evaluate the anonymity after the message is published. The experiment is running on a simulator [13] which based on a 5000×5000 square meters small town's road network. We generate 500, 1000, 2000, 4000 nodes as vehicles respectively and randomly and equably distribute on the road network. Those nodes move at a random initial velocity and are restricted on the road network. Those vehicles are attacked simulatively under two models: simple attack model and correlated attack model.

1) *Simple attack model.* The adversary randomly picks an observed peer. And then estimates an observed area Aob based on possible movement directions. It is a reachable circle area which depends on the max moving distance at a certain time. The vehicle's actual location at a future time must be in the area. Then the adversary will estimate where will the observed peer be after a observed time τ.
2) *Correlated attack model.* The correlated attack model is based on simple attack model. But it has an added condition. It presumes that the velocity of vehicles won't change dramatically during the observed time. It captures the initial and terminative velocity of every vehicle during the observed and eliminates the peers with a different velocity from observed peer.

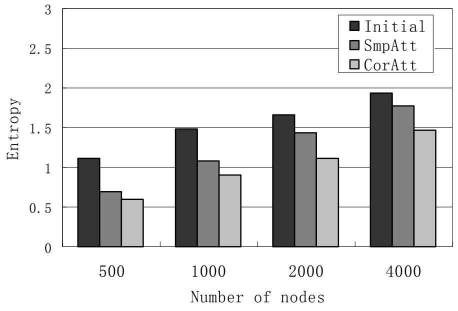


Figure 8. The anonymity under max velocity infer attack.

We calculate the entropy of adversary estimating the observed peer to evaluate the anonymity. Figure 8 shows the comparative results of initial status, simple attack model and correlated attack model. The result shows that the anonymity will enhance with the growth of peer density.

C. *Scalability*

The scale of VANET is large and dynamic. The scalability of VANET system is one of the most important aspects. Because of the restriction of wireless transmission range, the connectivity becomes an important indicator to evaluate the scalability of the system. We analysis two aspects: the scalability with nodes density and the scalability with anonymous path length.

Figure 9 shows the scalability with respect to the density of nodes. The connections are created in two ways: random mode and mobile-aware mode. Random mode randomly choice its next peer while mobile-aware mode consider the mobility of vehicles and choice the next peer with a better connectivity in future. The scalability will be better with the increase of node density for both of the two modes. But the mobile-aware mode has a better performance than random mode.

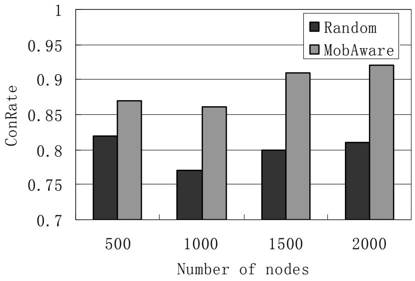


Figure 9. The scalability with respect to nodes density

Figure 10 shows the scalability with respect to path length. Theoretically, the path length can be no restricted. But because of the fast mobility of vehicles, the network topology changes dramatically. A longer path means more difficult to maintain the path. Figure 9 shows the remarkable influence of connectivity with the path length in different peer density.

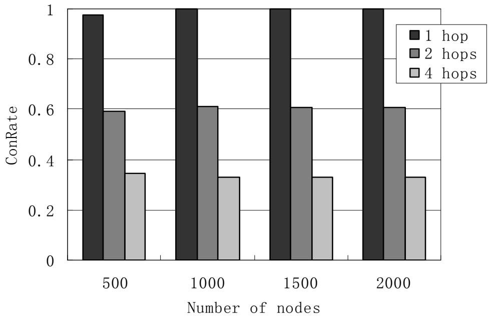


Figure 10. Figure 10 The scalability with respect to different path length

D. Authentication and Auditability

The system authenticates every vehicle before it joins to the VANET. They should be responsible for what they publish. Auditability is an important indictor to check fallacious messages. The peers in the VANET are forward peers and they only record where the message from and to in one hop. Figure 11 shows an example of forward tracing.

Though the peers in the path don't know who the originator of a message is, but they record its last hop and next hop. If we want to find a fallacious message's sender, we should gain all peers' records that on the anonymous path. Such a job can only be done by a special role with highest authority (e.g. police). If an accident is caused by a fallacious message, it can be audited later. The linkage protection can support the implement.

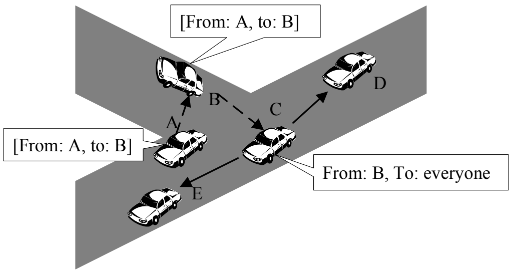


Figure 11. An example of forwarding trace

VI. Conclusion

This paper proposes a linkage protection algorithm for location privacy preserving. It achieves not only the unlinkability between identity and location, but also good anonymity, auditability and authentication. It also gives a new perspective in location privacy preserving. The analysis and experimental results can strongly support the proposition.

References

[1] John Krumm, "A survey of computational location privacy", Springer-Verlag Pers Ubiquit Comput (2009) 13:391–399

[2] HaoJin Zhu, "Security in service-oriented vehicular networks", IEEE Wireless Communications August 2009.

[3] Lars Kulik, "Privacy for Real-time Location-based Services", SIGSPATIAL Special 2009, Volume 1.

[4] John P. Baugh and Jinhua Guo, "Location Privacy in Mobile Computing Environments" UIC 2006, LNCS 4159, pp.936 – 945, 2006.

[5] Kido H, Yanagisawa Y and Satoh T, "An Anonymous Communication Technique Using Dummies for Location-based Service", Processing of IEEE International Conference on Pervasive Services. 2005: 88-97.

[6] Duckham M and Kulil L, "A formal model of obfuscation and negotiation for location privacy", Pervasive, 2005

[7] ChiYin Chow, Mohamed F. Mokbel and Xuan Liu, "A Peer-to-Peer Spatial Cloaking Algorithm for Anonymous Location based Services", ACMGIS'06, November 1011, 2006, Arlington, Virginia, USA.

[8] Gruteser M and Grunwald D, "Anonymous Usage of Location-based Services Through Spatial and Temporal Cloaking", Processing of the International Conference on Mobile Systems, Applications, and Services. MobiSys 2003 :163-168.

[9] Pfitamann A and Kohntopp M, "Anonymity, unobservability, and pseudonymity — a proposal for terminology", Designing Privacy Enhancing Technologies. Volume 2009 of Lecture Notes in Computer Science. Springer ,2001 :1-9.

[10] Beresford A R, Stajano F, "Location privacy in pervasive computing". IEEE Pervasive Computing, 2003: 46-55.

[11] Michael K. and Reiter, "Crowds: Anonymity for Web Transactions", ACM Transactions on Information and System Security, Vok. 1, No. 1, Novemeber 1998, Pages 66-92.

[12] K. Sampigethaya, L. Huang, M. Li, R. Poovendran, K. Matsuura and K. Sezaki, CARAVAN: providing location privacy for VANET, in: Proceedings of the Workshop on Embedded Security in Cars(escar)'05, 2005.

[13] T. Brinkhoff, "Generating Network-Based Moving Objects", In Proc. 12th International Conference on Scientific and Statistical Database Management, Berlin, Germany, 2000, pp. 253–255.