计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2014.20131074
51(11):2483-2492,2014

基于博弈分析的车辆感知网络节点轨迹隐私保护机制

何云华[1] 孙利民[1,2] 杨卫东[2,3] 李 志[2] 李 红[2]
[1](西安电子科技大学计算机学院 西安 710071)
[2](信息安全国家重点实验室(中国科学院信息工程研究所) 北京 100093)
[3](河南工业大学信息科学与工程学院 郑州 450001)
(heyunhua610@163.com)

Privacy Preserving for Node Trajectory in VSN: A Game-Theoretic Analysis Based Approach

He Yunhua[1], Sun Limin[1,2], Yang Weidong[2,3], Li Zhi[2], and Li Hong[2]
[1](*School of Computer Science and Technology, Xidian University, Xi'an* 710071)
[2](*State Key Laboratory of Information Security (Institute of Information Engineering, Chinese Academy of Sciences), Beijing* 100093)
[3](*College of Information Science and Engineering, Henan University of Technology, Zhengzhou* 450001)

Abstract By using crowdsourcing, vehicular sensor networks (VSN) are considered essential for achieving automatic and dynamic traffic information collection, and have created various fresh new business applications and services in our daily lives. However, the published trajectories that collected in VSNs raise significant privacy concerns. These existing methods, such as anonymization and cloaking techniques, though they are attractive for providing strong privacy guarantees, generally fail to satisfy the accuracy requirements of the trajectory data based applications. In addition, different attack strategies will result in quite different performance under various privacy preserving strategies. In order to address these challenges, we present a location privacy protection method, the DefenseGame algorithm. Given a set of trajectories and a probability density function for side information, the algorithm can assist the defender in selecting the optimal defense strategies by calculating the equilibriums in attack and defense games. In the attack and defense game, we use a game-theoretic model to capture the behavior of the adversary and defender, and we show the effectiveness of the two kinds of defense strategies in the adversary's inference attacks. Our experimental results indicate that the same defense strategy shows different performance for attack strategies and the proposed algorithm can help to obtain higher defender's utility compared with other approaches.

Key words vehicular sensor network (VSN); trace privacy; game theory; optimal defense strategies; equilibrium

摘 要 利用车辆移动"众包模式"为驾驶员提供实时路况信息和探索新的商业服务模式,是车载感知网络的新型应用之一.然而,车辆移动轨迹中的敏感信息易与用户身份相关联,存在隐私泄露问题.提出了

收稿日期:2013-07-25;**修回日期**:2013-10-15
基金项目:国家自然科学基金项目(61202099,61472418);国家"九七三"重点基础研究发展计划基金项目(2011CB302902);国家"八六三"高技术研究发展计划基金项目(2013AA011102);中国科学院先导科技专项课题(XDA06040100)
通信作者:孙利民(sunlimin@iie.ac.cn)

一种车载感知网络节点轨迹隐私保护算法,克服了匿名和隐藏技术缺乏数据真实性权衡的缺点,并考虑到多种攻击策略的影响.对于给定的一系列轨迹集,首先确定边信息概率分布,建立攻击者和防御者模型,通过求解攻防博弈中的纳什均衡选择最优的防御策略,并证明其有效性.在此基础上,折中轨迹数据真实性和隐私性,以最大化防御者效用为目标,提出轨迹隐私保护算法.实验结果表明,防御策略在不同的攻击策略下展现出不同的性能,算法所选择的防御策略优于其他的策略.

关键词 车载感知网络;轨迹隐私;博弈论;最优防御策略;均衡

中图法分类号 TP309.2

车辆感知网络(vehicular sensor network, VSN)通过收集动态交通信息,高效准确地感知城市交通分布,期望对人们的驾驶体验和城域交通管理产生深刻的变革,将毫无疑问地成为未来城域无线网络中重要的一部分[1].在 VSN 中,车辆的移动轨迹通常会定期发布.一方面,车辆的移动轨迹可以反映移动网络中节点的真实行为,有助于辅助设计和评估移动网络;另一方面,车辆的移动轨迹能够反映城市实时交通状况和分布,可用于城市道路规划.一些综合数据门户网站以及研究组织的网站已经提供了部分城市的车辆轨迹信息.

然而,轨迹中的时间和位置数据是很强的准标识符(quasi-identifiers),可以和其他类型的数据相关联[2],如住址区、工作地点、邮政编码等,因而很容易暴露真正的轨迹拥有者,对个人的人身安全、公司利益甚至国家安全造成威胁[3-4].另外,VSN 的开放性使得轨迹隐私更为严重[5].轨迹和位置信息不仅可能通过机会或预谋相遇得到,还可能从习惯、新闻、社交网络、博客等方式间接地推断出来[6-7].

轨迹隐私保护通常包括匿名(anonymization)[2,8]和隐藏(cloaking)技术[9-10].轨迹发布之前都需匿名化处理,车辆的真实 ID 通常被一个随机的标识符取代,然而轨迹中准标识符(时间和位置)可链接到其他数据库,来获取轨迹拥有者信息.虽然 k-匿名保证每个实体至少与其他 $k-1$ 个实体不可区分,但是轨迹中连续点之间具有依赖性,因此每条轨迹应视为整体进行匿名.轨迹整体匿名需要对轨迹中的位置进行修改,限制了轨迹数据的用途.另一类方法是隐藏技术,如减少记录数据和添加噪声数据等,这类方法减少了轨迹中连续点的依赖性,并在一定程度上保证了轨迹的真实性.然而,隐藏技术也存在一些缺陷:1)欠缺对隐私与轨迹真实性之间的权衡,不考虑最优化防御者的效用;2)很少考虑攻击策略对防御效果的影响,难以针对不同攻击策略给出最优防御策略.

为了克服以上缺点,我们采用博弈论来设计隐私保护机制.博弈论提供了很多工具和方法用于解决同一个系统中参与者之间的竞争和交互问题,并成功应用在经济学、决策理论、控制论等许多学科领域.博弈论非常适合于安全与隐私的分析、建模和决策过程.与纯优化方法相比,这种数学抽象方法有利于问题泛化和未来研究.

本文提出了一种轨迹隐私保护算法 DefenseGame.在给定一系列轨迹集和边信息概率分布的情况下,通过求解攻防博弈中的纳什均衡来选择最优的防御策略.博弈中,攻击者和防御者作为参与者,攻击者的策略是采用启发式的推断方法,防御者的策略是采用隐藏技术,攻击者和防御者的收益由轨迹的不确定性和攻防代价来计算.通过完全信息和不完全信息博弈分析,DefenseGame 算法最大化了防御者的效用.本文的主要贡献如下:

1) 对攻击者和防御者行为建模,给出了相应的攻击和防御策略,并证明了防御策略的有效性;

2) 提出的隐私保护算法 DefenseGame,权衡数据真实性和轨迹隐私,指导防御策略的选取;

3) 通过分析表明,不同的攻击策略下防御策略显示出不同的性能,DefenseGame 降低了攻击者的效用,最大化了防御者的效用.

1 相关工作

匿名(anonymization)[2,8,11-13]和隐藏(cloaking)技术[9-10,14]常用于保证节点轨迹隐私.轨迹发布之前需匿名化处理,车辆的真实 ID 通常被一个随机的标识符取代,然而轨迹中准标识符(时间和位置)可链接到其他数据库,来获取轨迹拥有者信息.Nergiz 等人采用 k-匿名的概念来达到轨迹匿名的目的,保证每条轨迹至少与其他 $k-1$ 个轨迹不可区分[2].然而,吴英杰等指出 k-匿名后的轨迹集仍然存在隐私威胁,并提出了基于$\langle k,\delta,\Delta\rangle$-匿名模型的匿名轨迹

集的隐私保护机制[11]. Chris 等人指出攻击者可以利用节点的少量位置快照(snapshot),标识出轨迹集中某一节点的轨迹[6]. Lu 等人提出一种社会热点匿名改变机制,来保证 VANET 中的位置隐私[13]. 他们将车辆聚集的点称为社会热点,如交叉路口或停车场,匿名集大小作为衡量隐私的指标,并且采用博弈论证明其方法的有效性. 这类方法或者没有考虑轨迹中连续点之间的依赖性,难以达到匿名的目的,或者需要对轨迹中的位置进行修改,限制了轨迹数据的用途.

另一类方法是隐藏技术,如减少记录数据和添加噪声数据等,这类方法减少轨迹中连续点的依赖性,并在一定程度上保证轨迹的真实性. Tang 等人提出了一些隐私基准,建议抽样频率(即探测位置更新的频率)应该限制到更大时间间隔上[14]. Hoh 等人提出了一种时间混淆标准来定义位置数据集中的隐私,并提出一种不确定发现轨迹混淆算法,隐藏数据集中的某些抽样点,为车辆轨迹提供时间上的混淆[10]. 然而,隐藏技术很少考虑在隐私与轨迹真实性之间作出权衡,几乎没有考虑攻击策略对防御效果的影响,也就不能针对不同攻击策略给出最优防御策略,难以保证达到所需的隐私要求.

大量的研究人员使用博弈论来研究移动网络中的安全与隐私问题[15-18]. Raya 等人研究了短暂网络中撤销博弈问题,其中采用参与者合作来决定是否撤销潜在的恶意参与者[15]. Reidt 等人采用博弈论方法设计了一种 ad hoc 网络中的分布式健壮动态撤销机制,激励诚实节点对恶意节点进行撤销[16]. Alpcan 等人采用零和博弈、模糊博弈和“虚拟游戏”来建模 VANET 中攻击者和防御措施之间交互过程[17],其中路段重要性由网络的核心性来度量. Freudiger 等人采用博弈论研究 ad hoc 网络中的位置隐私问题,分析了网络中移动节点的非合作行为,建立了 n 个参与者之间的完全信息博弈和不完全信息博弈,并证明了纳什均衡的存在性[18]. 本文则采用博弈论来优化设计轨迹隐私保护机制.

2 模型假定

VSN 的网络模型如图 1 所示,车辆采用专用短程通信(dedicated short range communications, DSRC)技术,将包含位置、时间戳的感知信息传递到路边单元(roadside unit, RSU),RSU 通过有线网络上传到交管中心,从而交管中心能够高效准确地感知城市交通分布[1]. VSN 定期公布轨迹集合,包含一系列的时间和所对应移动车辆的位置信息. 这些轨迹信息由车辆 GPS 设备以众包模式收集,上传给交管中心. 这些轨迹都是匿名的,因为车辆的真实 ID 被随机标识所取代. 车辆的真实 ID 与随机标识没有任何联系,但是一个真实 ID 总是映射到同一个标识.

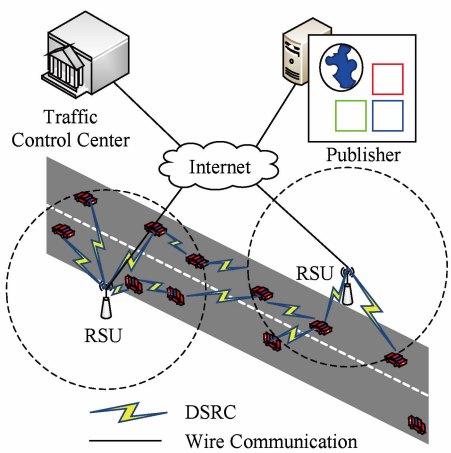


Fig. 1 Architecture of vehicular sensor network.

图 1 VSN 的网络模型

2.1 攻击者模型

攻击者的目的是从匿名轨迹集 Tr 中确定一个或多个车辆(知道其真实 ID)的完整历史轨迹. 假定攻击者可以收集关于一个或多个车辆轨迹的边信息(side information),每一份边信息给出了时刻 t 车辆 $i\in U$ 的位置信息 $r_i(t)$,该信息可能是不准确的. 实际上,边信息可以通过以下方式获得:1)节点在公共场所是可观察的,即攻击者可以直接通过机会或预谋相遇来获取受害节点的边信息[8];2)节点可能自愿或无意地泄露自己的位置信息[7],例如 Alice 与 Bob 闲谈中透露“下午 3 点 Alice 在公司”.

设$\{o_k\}$为节点轨迹发布的可观察的抽样时间,它是等距分布的,$\{t_k\}$为边信息所揭示受害节点的时间实例,那么边信息推断的时间实例位于 2 个连续抽样时间之间,即$\{t_k\mid k=1,2,\cdots\}\not\subset\{o_k\mid k=1,2,\cdots\}$,且对于每个 t_k 存在 $\tilde{k}$,使得:

$$t_k=\lambda_k o_{\tilde{k}}+(1-\lambda_k)o_{\tilde{k}+1},\ 0\leqslant\lambda_k<1,$$

$\lambda_k=0$ 时,正好位于可观察的抽样点.

由于边信息通常包含噪声,攻击者需要采用贝叶斯推断. 攻击者进行推断的目标是,在给定 r_i 的情况下,查找轨迹 $Tr_j\in Tr$ 以获得最优的匹配结果.

这个过程的形式化描述如下:给定 $r_i=\{r_i(t_k)\mid k=1,2,\cdots\}$,计算:

$$Pr(Tr_j\mid r_i)=\frac{Pr(Tr_j,r_i(t_k))}{Pr(r_i(t_k))}=\frac{Pr(r_i(t_k)\mid Tr_j)Pr(Tr_j)}{\sum_{h=1}^{N}Pr(r_i(t_k)\mid Tr_h)Pr(Tr_h)},\quad i\in U,k=1,2,\cdots. \tag{1}$$

攻击者的目标是寻找 j,使得式(1)值最大,分母为常数.在贝叶斯推断过程中,攻击者也可以利用一些常识,如节点的移动限制,节点会受到道路的拓扑结构和节点的移动规律的影响.

2.2 防御者模型

在 VSN 中,防御者可以是交管中心或权威机构,其目标为保证某一轨迹以较低概率从发布轨迹中确定.主要假设如下:1)防御者知道攻击者可能采用的攻击策略,但不知道具体采取哪一种攻击策略;2)防御者对攻击者掌握的边信息有一定(完全或部分)的了解;3)防御策略会带来一定的损失.

设$\{t_k\}$为边信息所揭示受害节点的时间实例,Ω为防御者认为攻击者已掌握的边信息,那么防御者的目标可表示为,在给定 Ω 的情况下,使得

$$Pr(Tr_i\mid\Omega)=\frac{Pr(\Omega\mid Tr_i(t_k))Pr(Tr_i(t_k))}{Pr(\Omega)} \tag{2}$$

较小,其中 $k=1,2,\cdots$.本文采用的防御策略是增加虚假路径集和减少抽样时间点,将在 3.2.2 节详细介绍.

3 攻防博弈分析模型

3.1 博弈模型

攻防博弈 G 可定义为三元组(P,S,U),其中 P 是参与者集合,S 是策略集,U 是收益函数集[19].当防御者知道攻击者所掌握的边信息时,建立完全信息博弈,此时攻击者的收益为

$$u_1(s_1,s_2)=-\sum_i p_i\mathrm{lb}(p_i)-\gamma_1=H-\gamma_1,$$

防御者的收益为

$$u_2(s_1,s_2)=-\sum_i p_i\mathrm{lb}(p_i)-\gamma_2=H-\gamma_2.$$

其中,p_i 代表目标节点被识别的概率;$H=-\sum_i p_i\mathrm{lb}(p_i)$为轨迹的不确定性,$H$ 越小攻击者确定目标轨迹的概率越高,目标节点的轨迹隐私越低;γ_1,γ_2 分别表示攻击和防御代价,这里为了最大化攻击者的能力,将 γ_1 设为 0.γ_2 包括防御措施的代价,以及它对数据真实性造成的影响.

然而,通常防御者不完全知道攻击者所掌握的边信息,因此我们考虑采用 Harsanyi[20] 的方法,引入一个新参与者"自然"来建立不完全信息博弈.设防御者的类型为 θ,服从先验分布 p,收益函数可表示为 $u_2(s_1,s_2(\theta))$.

在完全信息博弈中,我们将均衡定义如下:

定义 1. 策略 s^* 是纳什均衡,如果每个参与者 i 有:

$$u_i(s_i^*,s_{-i}^*)\geqslant u_i(s_i,s_{-i}^*),\ \forall s_i\in S_i.$$

换句话说,在纳什均衡中,没有参与者可以通过改变策略来增加收益.

在不完全信息博弈中,我们采用贝叶斯纳什均衡的概念[19],如定义 2 所示:

定义 2. 策略组合 $s^*=(s_i^*(\theta_i),s_{-i}^*(\theta_{-i}))$是贝叶斯纯策略均衡,如果对每个参与者 i 有:

$$s^*\in\arg\max_{s_i\in S_i}\sum_{\theta_{-i}}p(\theta_{-i})\times u_i(s_i,s_{-i}^*(\theta_{-i})),\forall\theta_i.$$

3.2 攻防策略

3.2.1 攻击策略

由于选取受害节点没有任何可参考的信息,因此我们设定受害者服从均匀分布:$Pr(Tr_h)=1/N$,$h=1,2,\cdots,N$.因此,式(1)的解如下:

$$\max_{i=1,2,\cdots,N}Pr(r_i(t_k)\mid Tr_j),k=1,2,\cdots. \tag{3}$$

假设节点移动遵循 Markov 模型,式(3)可以写成如下形式:

$$Pr(r_i(t_k)\mid Tr_j)=Pr(r_i(t_k)\mid Tr_j(\{o_k\mid k=1,2,\cdots\}))=\frac{\prod_k[Pr(Tr_j(o_{k+1})\mid R(t_k))\times Pr(r_i(t_k)\mid Tr_j(o_k))]}{\prod_k Pr(Tr_j(o_{k+1})\mid Tr_j(o_k))}, \tag{4}$$

其中,$t_k=\lambda_k o_k+(1-\lambda_k)o_{k+1}$.当 λ_k 为常数时,节点移动的 Markov 过程是平稳的.

上面给出了贝叶斯推断的严格数学表达式,可根据它导出确定受害节点轨迹的一些简单启发式的方法.接下来,我们将考虑 4 种启发式算法[6].

为了方便阐述,将时刻 t_k 车辆在位置 x 的概率定义为 $P_j:\Theta\times\{t_k\mid t=1,2,\cdots\}\to R_+$,详细如下:

$$P_j(x,t_k)=\frac{Pr(Tr_j(o_{k+1})\mid r_i(t_k))\times Pr(r_i(t_k)\mid Tr_j(o_k))}{Pr(Tr_j(o_{k+1})\mid Tr_j(o_k))},$$

其中,$x=\lambda_k Tr_j(o_k)+(1-\lambda_k)Tr_j(o_{k+1})$,$o_k\leqslant t_k<o_{k+1}$,$\Theta$ 为所有单元格(cell)位置 ID 集合.我们得到:

1) 最大似然函数方法(MLE). 轨迹 j 的相似度值由

$$\prod_k \Big(\sum_{x\in\Theta} P_j(x,t_k)Pr_Z(r_i(t_k)-x)\Big) \tag{5}$$

给出,最大相似度值的轨迹将是受害节点的轨迹. 噪声 Z_k 是独立同分布的,服从给定分布 Pr_Z.

2) 最小二(平方)乘法(MSQ). 轨迹 j 的相似度值由

$$\sum_k \Big(\sum_{x\in\Theta} P_j(x,t_k)\mid r_i(t_k)-x\mid^2\Big) \tag{6}$$

给出,最小负相似度值的轨迹将是受害节点的轨迹.

3) 基本方法(basic approach, BAS). 假定噪声的均值为 0,标准差为 σ. 攻击者利用边信息计算轨迹 j 的相似度值:

$$\sum_k \Big(\sum_{x\in\Theta} P_j(x,t_k)\times I_{2\sigma}(x,r_i(t_k))\Big), \tag{7}$$

其中,当$|x-y|\leqslant 2\sigma$ 时,$I_{2\sigma}(x,y)=1$,否则 $I_{2\sigma}(x,y)=0$.

4) 加权指数方法(EXP). 类似于 BAS,攻击者计算轨迹 j 的最大相似度值:

$$\sum_k \Big(\sum_{x\in\Theta} \frac{P_j(x,t_k)}{w(r_i(t_k))}\exp\Big(-\frac{1}{C}\mid x-r_i(t_k)\mid\Big)\Big), \tag{8}$$

其中,$w(r_i(t_k))$是分配给 $r_i(t_k)$的权值,C 为常数.

3.2.2 防御策略

常见的防御策略有匿名技术和隐藏技术. 匿名技术将用户的真实身份隐藏,其实现方式有多种,如用户使用一个或多个假名[1,21]或者一组用户共用同一个 ID[22]. 然而在 VSN 中限定车辆的真实 ID 总映射到同一个标识,所以本文将不对匿名技术展开讨论. 隐藏技术也可用来保证用户的轨迹隐私,如添加噪声数据或减少记录数据. VSN 中采用隐藏技术将影响所获取信息的准确度,因此我们需要在数据真实性和轨迹隐私之间作出权衡.

设 Tr 为节点的轨迹集,$\{o_1,o_2,\cdots,o_k,\cdots\}$为可观察的抽样时间序列,防御策略可表述如下:

策略 1($D1$). 向 Tr 添加 m 条轨迹组成新的轨迹集 Tr',使得 $Tr'\supset Tr$,且满足$|Tr'|=|Tr|+m$.

策略 2($D2$). 删除可观察序列$\{o_1,o_2,\cdots,o_k,\cdots\}$中的 n 个抽样点,组成新的序列$\{o'_1,o'_2,\cdots,o'_k,\cdots\}$,使得$\{s_1,s_2,\cdots,s_k,\cdots\}\supset\{o'_1,o'_2,\cdots,o'_k,\cdots\}$.

定理 1. $D1$ 中节点所能提供的隐私增加.

证明. 执行 $D1$ 前,节点所能提供的隐私满足:

$$H=-\sum_{i=1}^{|Tr|} p_i\mathrm{lb}(p_i)\leqslant \sum_{i=1}^{|Tr|}\frac{1}{|Tr|}\mathrm{lb}(|Tr|)=\mathrm{lb}(|Tr|).$$

执行 $D1$ 后,节点所能提供的隐私满足:

$$H=-\sum_{i=1}^{|Tr'|} p_i\mathrm{lb}(p_i)\leqslant \sum_{i=1}^{|Tr'|}\frac{1}{|Tr'|}\mathrm{lb}(|Tr'|)=\mathrm{lb}(|Tr'|)=\mathrm{lb}(|Tr|+m).$$

由于 $\mathrm{lb}(|Tr|+m)>\mathrm{lb}(|Tr|)$,因此节点所能提供的隐私增加.

定理 2. $D2$ 降低攻击者识别轨迹的概率.

证明.

$$Pr(r_i(t_k)\mid Tr_j)=Pr(r_i(t_k)\mid Tr_j(\{o_1,o_2,\cdots,o_k,\cdots\}))\geqslant Pr(r_i(t_k)\mid Tr_j(\{o'_1,o'_2,\cdots,o'_k,\cdots\}))=Pr(r_i(t_k)\mid Tr'_j),$$

其中,$\{o_1,o_2,\cdots,o_k,\cdots\}\subset\{o'_1,o'_2,\cdots,o'_k,\cdots\}$. 由于 $Pr(r_i(t_k)\mid Tr_j)\geqslant Pr(r_i(t_k)\mid Tr_j')$,$D2$ 降低轨迹被攻击者识别的概率. 证毕.

然而,防御措施会对数据真实性造成影响. 增加虚假轨迹数越多,删除的抽样时间点越多,数据越不真实. 因此,我们假定当增加虚假轨迹数越多,删除的抽样时间点越多,防御开销 γ 越大.

设在 $D1$ 下,防御代价 γ 满足以下关系式:

$$\gamma=\alpha\frac{m}{|Trace|},$$

其中 α 为比例参数,$|Trace|$ 为节点轨迹总数,m 为添加轨迹的条数.

设在 $D2$ 下,防御代价 γ 满足以下关系式:

$$\gamma=\beta\frac{n}{N},$$

其中,β 为比例参数,N 为抽样时间点的总数,n 为删除抽样点的个数.

4 隐私保护机制

本节提出了一种节点轨迹隐私保护算法 Defense-Game,该算法能够最大化防御者的效用. 在给定一系列轨迹集和边信息概率分布的情况下,通过求解攻防博弈中的纳什均衡来选择最优的防御策略.

由于攻击者所获取的边信息未知,因此假定边信息中包含的$\langle time,position\rangle$对的个数服从概率分布 p,c_l 表示边信息包含 l 个$\langle time,position\rangle$对. 考察不完全信息博弈,我们观察到,当博弈达到纳什均衡时,防御者通过改变策略不能再增加其收益,即此时防御者的收益达到最大.

算法 1. DefenseGame 算法.

输入:轨迹集$\{Tr_i \mid i=1,2,\cdots,|Tr|\}$和边信息分布 p;

输出:更新后的轨迹集$\{Tr'_i\}$.

① If 防御者选择 $D1$ then

② 计算 s_2^*:

$$s_2^* \in \arg\max_{s_2 \in D1}\{p_1 u_2(s_1, s_2(c_1)) + p_2 u_2(s_1, s_2(c_2)) + \cdots + p_l u_2(s_1, s_2(c_l))\};$$

③ 插入 $m=s_2^*$ 条虚拟轨迹到$\{Tr_i \mid i=1,2,\cdots,|Tr|\}$;

④ Return 新轨迹集 $\{Tr'_i \mid i=1,2,\cdots,|Tr|+m\}$;

⑤ Else

⑥ 计算 s_2^*:

$$s_2^* \in \arg\max_{s_2 \in D2}\{p_1 u_2(s_1, s_2(c_1)) + p_2 u_2(s_1, s_2(c_2)) + \cdots + p_l u_2(s_1, s_2(c_l))\};$$

⑦ 从$\{o_1, o_2, \cdots, o_k, \cdots\}$中删除 $n=s_2^*$ 个抽样点;

⑧ Return 新轨迹集$\{Tr'_i\}=\{Tr_i(o'_k)\}$;

⑨ End if

算法 1 给出了 DefenseGame 算法的基本思想.该算法描述了分别在 $D1$,$D2$ 下计算的纳什均衡,然后根据纳什均衡来选择在 $D1$,$D2$ 下的最优防御策略.它以轨迹集、边信息概率分布 p 为输入,以更新后的轨迹集为输出,这些轨迹集可以发布,有一定的隐私保护作用.

DefenseGame 算法中的纳什均衡依赖于攻击者和防御者的收益函数.攻击者的收益函数仅由轨迹的不确定性计算,防御者的收益函数由轨迹不确定性和数据的真实性计算.这样计算是合理的,因为攻击者通常是强大的,不考虑攻击代价,恰好最大化攻击者的能力;防御者是理性的,通常要考虑其代价与利益,轨迹隐私和数据真实性正是防御者主要考虑的 2 个方面.既然计算更贴切现实情况,DefenseGame 算法对选取合理的防御策略是有指导意义的.

5 实验与分析

旧金山出租车的真实轨迹被用于本文的博弈论分析,实验的主要参数如表 1 所示.我们分别进行了完全信息和不完全信息博弈分析.

Table 1 Parameters of Experiment

表 1 实验参数

Parameters	San Francisco Cabs
$Latitude_{\min}$	37.05
$Latitude_{\max}$	38.00
$Longitude_{\min}$	−122.86
$Longitude_{\max}$	−122.00
Cells	8 170
Active Cells	3 997
$Timestamp_{\min}$	2008-05-17 03:00:04
$Timestamp_{\max}$	2008-06-10 02:25:34

攻击者有 4 种攻击策略:MLE,MSQ,BAS,EXP,其中边信息是一条轨迹的随机抽样点加入高斯噪声得到的.防御者采用 2 种防御策略:$D1$,$D2$.在 $D1$ 中,设置 $m=0$,$|Trace|/8$,$|Trace|/4$,$|Trace|/2$,其中$|Trace|$为轨迹总数;在 $D2$ 中,设置 $n=0$,$N/2$,$3N/4$,$7N/8$,其中 N 为抽样时间点总数.

5.1 完全信息博弈分析

防御者完全知道攻击者获取的边信息,即防御者知道攻击者的收益,这里边信息含有 18 个$\langle time, position\rangle$对.由于增加的轨迹越多,删除的抽样时间点越多,防御代价越大,因此我们不妨假定 $\gamma^1=\alpha/8$,$\gamma^2=\alpha/4$,$\gamma^3=\alpha/2$,$\gamma^4=\beta/2$,$\gamma^5=3\beta/4$,$\gamma^6=7\beta/8$,其中 α,β 为比例系数.完全信息博弈的策略式如图 2 所示.可以观察到,攻击者采取 BAS 时防御效果最优;攻击者采用 MLE 时防御效果最差,因此攻击策略对防御效果产生了影响.

The Defender Chooses $D1$

	$m=0$	$m=\frac{\lvert Trace\rvert}{8}$	$m=\frac{\lvert Trace\rvert}{4}$	$m=\frac{\lvert Trace\rvert}{2}$
MLE	0.37,0.37	0.43,0.43−γ^1	0.57,0.57−γ^2	0.78,0.78−γ^3
MSQ	0.38,0.38	0.44,0.44−γ^1	0.59,0.59−γ^2	0.79,0.79−γ^3
BAS	0.80,0.80	0.97,0.97−γ^1	1.10,1.10−γ^2	1.35,1.35−γ^3
EXP	0.67,0.67	0.80,0.80−γ^1	0.96,0.96−γ^2	1.18,1.18−γ^3

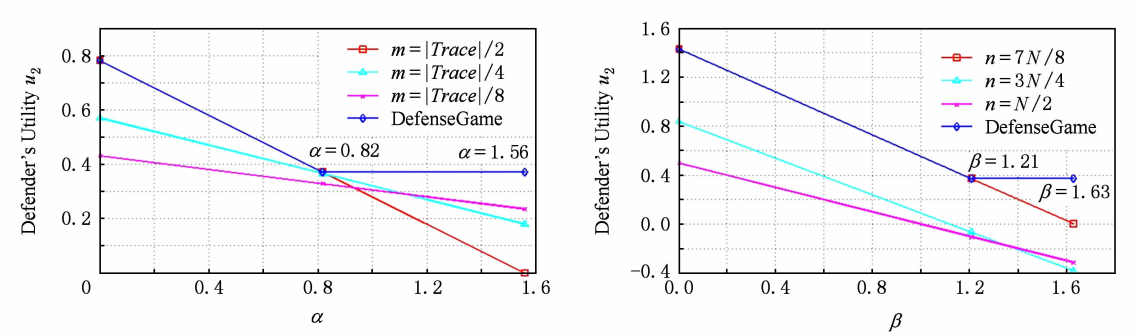
The Defender Chooses $D2$

	$n=0$	$n=N/2$	$n=3N/4$	$n=7N/8$
MLE	0.37,0.37	0.50,0.50−γ^4	0.84,0.84−γ^5	1.43,1.43−γ^6
MSQ	0.38,0.38	0.58,0.58−γ^4	1.11,1.11−γ^5	2.37,2.37−γ^6
BAS	0.80,0.80	1.04,1.04−γ^4	1.58,1.58−γ^5	2.78,2.78−γ^6
EXP	0.76,0.76	0.92,0.92−γ^4	1.43,1.43−γ^5	2.44,2.44−γ^6

Fig. 2 Strategic form for complete information game.

图 2 完全信息博弈的策略式

当 $0\leqslant\alpha\leqslant 0.82$ 时,DefenseGame 算法将 $m=|Trace|/2$ 作为其防御策略,熵值 H 从 0.37 增加到 0.78;当 $\alpha\geqslant 0.82$ 时,将 $m=0$ 作为防御策略,即不采用任何防御措施. 当 $0\leqslant\beta\leqslant 1.21$ 时,算法将 $n=7N/8$ 作为其防御策略,熵值 H 从 0.37 增加到 1.43;当 $\beta\geqslant 1.21$ 时,算法将 $n=0$ 作为防御策略. 图 3 比较了 DefenseGame 算法与其他策略. 可以观察到,DefenseGame 算法优于其他防御策略,最大化了防御者的效用,这是因为达到纳什均衡时参与者无法通过改变策略来增加其效用.



(a) $D1$ has been chosen (b) $D2$ has been chosen

Fig. 3 Defender's utility comparison of DefenseGame with other strategies.

图 3 DefenseGame 选取的策略与其他策略的比较

5.2 不完全信息博弈分析

通常,防御者不能完全知道攻击者所掌握的边信息,因而 harsanyi[20] 将"自然"作为参与者,把不完全信息博弈转化为完全但不完美的信息博弈. 如图 4

	$m=0$	$m=\frac{\lvert Trace\rvert}{8}$	$m=\frac{\lvert Trace\rvert}{4}$	$m=\frac{\lvert Trace\rvert}{2}$
MLE	1.01,1.01	1.09,1.09$-\gamma^1$	1.26,1.26$-\gamma^2$	1.59,1.59$-\gamma^3$
MSQ	1.02,1.02	1.11,1.11$-\gamma^1$	1.27,1.27$-\gamma^2$	1.61,1.61$-\gamma^3$
BAS	1.56,1.56	1.69,1.69$-\gamma^1$	1.89,1.89$-\gamma^2$	2.18,2.18$-\gamma^3$
EXP	1.45,1.45	1.56,1.56$-\gamma^1$	1.78,1.78$-\gamma^2$	2.05,2.05$-\gamma^3$

The Defender Chooses $D1$

	$n=0$	$n=N/2$	$n=3N/4$	$n=7N/8$
MLE	1.01,1.01	1.15,1.15$-\gamma^4$	1.61,1.61$-\gamma^5$	2.50,2.50$-\gamma^6$
MSQ	1.02,1.02	1.26,1.26$-\gamma^4$	1.84,1.84$-\gamma^5$	3.17,3.17$-\gamma^6$
BAS	1.56,1.56	1.86,1.86$-\gamma^4$	2.45,2.45$-\gamma^5$	3.65,3.65$-\gamma^6$
EXP	1.45,1.45	1.75,1.75$-\gamma^4$	2.30,2.30$-\gamma^5$	3.25,3.25$-\gamma^6$

The Defender Chooses $D2$

(a) The probability that side information contains 10 pairs of $\langle time, location\rangle$ is p_1

	$m=0$	$m=\frac{\lvert Trace\rvert}{8}$	$m=\frac{\lvert Trace\rvert}{4}$	$m=\frac{\lvert Trace\rvert}{2}$
MLE	0.51,0.51	0.64,0.64$-\gamma^1$	0.79,0.79$-\gamma^2$	1.01,1.01$-\gamma^3$
MSQ	0.53,0.53	0.65,0.65$-\gamma^1$	0.81,0.81$-\gamma^2$	1.03,1.03$-\gamma^3$
BAS	1.03,1.03	1.18,1.18$-\gamma^1$	1.36,1.36$-\gamma^2$	1.56,1.56$-\gamma^3$
EXP	0.89,0.89	1.03,1.03$-\gamma^1$	1.20,1.20$-\gamma^2$	1.45,1.45$-\gamma^3$

The Defender Chooses $D1$

	$n=0$	$n=N/2$	$n=3N/4$	$n=7N/8$
MLE	0.51,0.51	0.67,0.67$-\gamma^4$	1.04,1.04$-\gamma^5$	1.85,1.85$-\gamma^6$
MSQ	0.53,0.53	0.74,0.74$-\gamma^4$	1.29,1.29$-\gamma^5$	2.58,2.58$-\gamma^6$
BAS	1.03,1.03	1.27,1.27$-\gamma^4$	1.80,1.80$-\gamma^5$	2.97,2.97$-\gamma^6$
EXP	0.89,0.89	1.16,1.16$-\gamma^4$	1.67,1.67$-\gamma^5$	2.68,2.68$-\gamma^6$

The Defender Chooses $D2$

(b) The probability that side information contains 15 pairs of $\langle time, location\rangle$ is p_2

	$m=0$	$m=\frac{\lvert Trace\rvert}{8}$	$m=\frac{\lvert Trace\rvert}{4}$	$m=\frac{\lvert Trace\rvert}{2}$
MLE	0.27,0.27	0.36,0.36$-\gamma^1$	0.43,0.43$-\gamma^2$	0.64,0.64$-\gamma^3$
MSQ	0.28,0.28	0.37,0.37$-\gamma^1$	0.45,0.45$-\gamma^2$	0.67,0.67$-\gamma^3$
BAS	0.69,0.69	0.82,0.82$-\gamma^1$	0.96,0.96$-\gamma^2$	1.18,1.18$-\gamma^3$
EXP	0.56,0.56	0.69,0.69$-\gamma^1$	0.81,0.81$-\gamma^2$	1.03,1.03$-\gamma^3$

The Defender Chooses $D1$

	$n=0$	$n=N/2$	$n=3N/4$	$n=7N/8$
MLE	0.27,0.27	0.39,0.39$-\gamma^4$	0.70,0.70$-\gamma^5$	1.45,1.45$-\gamma^6$
MSQ	0.28,0.28	0.46,0.46$-\gamma^4$	0.95,0.95$-\gamma^5$	2.18,2.18$-\gamma^6$
BAS	0.69,0.69	0.92,0.92$-\gamma^4$	1.43,1.43$-\gamma^5$	2.56,2.56$-\gamma^6$
EXP	0.56,0.56	0.80,0.80$-\gamma^4$	1.28,1.28$-\gamma^5$	2.29,2.29$-\gamma^6$

The Defender Chooses $D2$

(c) The probability that side information contains 20 pairs of $\langle time, location\rangle$ is $1-p_1-p_2$

Fig. 4 Strategic form for incomplete information game.

图 4 不完全信息博弈的策略式

所示,我们假设“自然”选择“边信息包含 10 个抽样点”的概率为 p_1(如图 4(a)所示),“边信息包含 15 个抽样点”的概率为 p_2(如图 4(b)所示),“边信息包含 20 个抽样点”的概率为 $1-p_1-p_2$(如图 4(c)所示). 可以观察到,MLE 总是攻击者的严格占优策略,此时防御效果最差.

我们考察了不同概率分布下的 DefenseGame 算法与其他策略的性能,如图 5 所示:

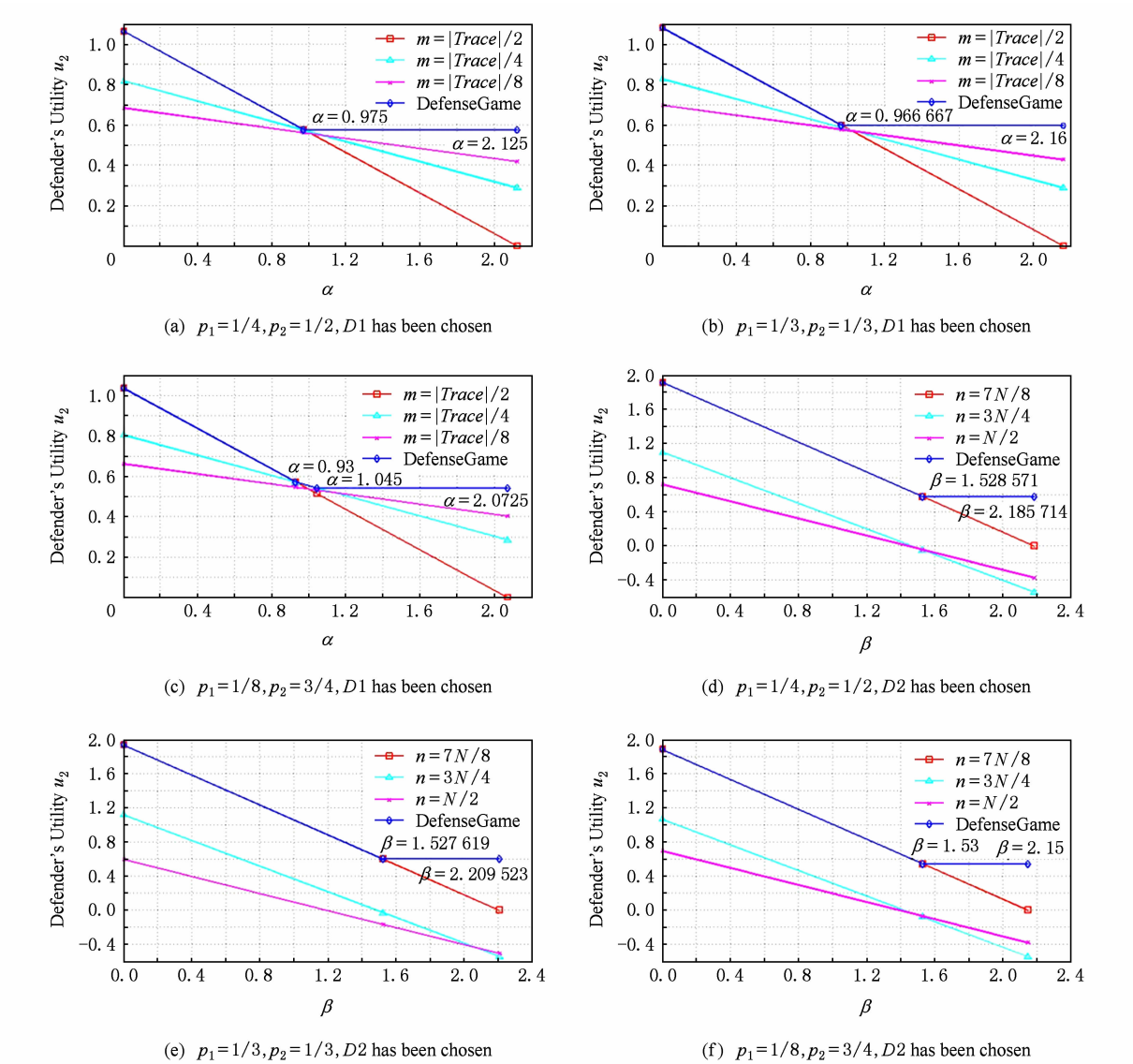


Fig. 5 Defender's utility comparison of DefenseGame with other strategies at different probability distrubitions.

图 5 DefenseGame 在不同概率分布下选取的策略与其他策略的比较

当 $p_1=1/4, p_2=1/2$ 时,如果 $0\leqslant\alpha\leqslant0.975$,算法选择 $m=|Trace|/2$,如果 $\alpha\geqslant0.975$,选择 $m=0$;如果 $0\leqslant\beta\leqslant1.539$,算法选择 $n=7N/8$,如果 $\beta\geqslant1.539$,选择 $n=0$. 当 $p_1=1/3, p_2=1/3$ 时,如果 $0\leqslant\alpha\leqslant0.967$,算法选择 $m=|Trace|/2$,如果 $\alpha\geqslant0.967$,选择 $m=0$;如果 $0\leqslant\beta\leqslant1.528$,算法选择 $n=7N/8$,如果 $\beta\geqslant1.528$,选择 $n=0$. 当 $p_1=1/8, p_2=3/4$ 时,如果 $0\leqslant\alpha\leqslant0.93$,算法选择 $m=|Trace|/2$,如果 $0.93\leqslant\alpha\leqslant1.045$,算法选择 $m=|Trace|/4$,如果 $\alpha\geqslant1.045$,选择 $m=0$;如果 $0\leqslant\beta\leqslant1.528$,算法选择 $n=7N/8$,如果 $\beta\geqslant1.528$,选择 $n=0$. 总体上看,DefenseGame 算法始终优于其他策略,这是因为 DefenseGame 根据纳什均衡选取防御策略.

5.3 实验分析结论

完全信息和不完全信息博弈均表明,防御策略的性能依赖于攻击者采用的策略,$D2$ 的性能均略优于 $D1$. DefenseGame 算法选取的策略比其他策略性能优,在降低攻击者的效用的同时,也能兼顾数据的

真实性.

6 总 结

本文基于博弈论分析,提出了一种车载感知网络节点轨迹隐私保护算法 DefenseGame.我们首先对攻击者和防御者行为建模,给出相应的攻击和防御策略,并证明了防御策略的有效性;然后,提出基于博弈论分析的节点轨迹隐私保护算法 DefenseGame,能够最大化防御者的效用;最后,通过实验验证了 DefenseGame 算法的有效性.DefenseGame 算法选择的策略优于其他策略,不仅能够降低攻击者的效用,而且可以最大化防御者的效用.我们还发现,$D2$ 略优于 $D1$,它们的性能均依赖于攻击策略.

参 考 文 献

[1] Zhang Chenxi, Lu Rongxing, Lin Xiaodong, et al. An efficient identity-based batch verification scheme for vehicular sensor networks [C] //Proc of the 27th Annual IEEE Int Conf on Computer Communications. Piscataway, NJ: IEEE, 2008: 816-824

[2] Nergiz M, Atzori M, Saygın Y, et al. Towards trajectory anonymization: A generalization-based approach [J]. Journal Trans on Data Privacy, 2009, 2(1): 47-75

[3] Lin Xiaodong, Lu Rongxing, Zhang Chenxi, et al. Security in vehicular ad hoc networks [J]. IEEE Communications Magazine, 2008, 46(4): 88-95

[4] Razzaque M, Ahmad S, Seyed M. Security and Privacy in Vehicular Ad-Hoc Networks: Survey and the Road Ahead [M]. Berlin: Springer, 2013: 107-132

[5] Mershad K, Artail H. A framework for secure and efficient data acquisition in vehicular ad hoc networks [J]. IEEE Trans on Vehicular Technology, 2013, 62(2): 536-551

[6] Chris M, David Y, Nung Y. Privacy vulnerability of published anonymous mobility traces [C] //Proc of ACM MobiCom'10. New York: ACM, 2010: 185-196

[7] Narayanan A, Shmatikov V. Robust de-anonymization of large sparse datasets [C] //Proc of the 29th IEEE Symp on Security and Privacy. Piscataway, NJ: IEEE, 2008: 111-125

[8] Huo Zheng, Meng Xiaofeng. A survey of trajectory privacy-preserving techniques [J]. Chinese Journal of Computers, 2011, 34(10): 1820-1830 (in Chinese)
(霍峥, 孟小峰. 轨迹隐私保护技术研究[J]. 计算机学报, 2011, 34(10): 1820-1830)

[9] Gruteser M, Grunwald D. Anonymous usage of location-based services through spatial and temporal cloaking [C] // Proc of ACM MobiSys'03. New York: ACM, 2003: 31-42

[10] Hoh B, Gruteser M, Xiong Hui, et al. Preserving privacy in gps traces via uncertainty-aware path cloaking [C] //Proc of the 14th ACM Conf on Computer and Communications Security. New York: ACM, 2007: 161-171

[11] Wu Yingjie, Tang Qingming, Ni Weiwei, et al. A clustering hybrid based algorithm for privacy preserving trajectory data publishing [J]. Journal of Computer Research and Development, 2013, 50(3): 578-593 (in Chinese)
(吴英杰, 唐庆明, 倪巍伟, 等. 基于聚类杂交的隐私保护轨迹数据发布算法[J]. 计算机研究与发展, 2013, 50(3): 578-593)

[12] IEEE Vehicular Society. IEEE Std 1609. 2-2013: IEEE Standard for Wireless Access in Vehicular Environments-Security Services for Applications and Management Messages [S]. New York: IEEE-SA Standards Board, 2013

[13] Lu Rongxing, Lin Xiaodong, Luan T, et al. Pseudonym changing at social spots: An effective strategy for location privacy in VANETs [J]. IEEE Trans on Vehicular Technology, 2012, 61(1): 86-96

[14] Tang P, Keyani P, Fogarty J, et al. Putting people in their place: An anonymous and privacy-sensitive approach to collecting sensed data in location-based applications [C] // Proc of the SIGCHI Conf on Human Factors in Computing Systems. New York: ACM, 2006: 93-102

[15] Raya M, Manshaei M, Felegyhazi M, et al. Revocation games in ephemeral networks [C] //Proc of the 15th ACM Conf on Computer and Communications Security. New York: ACM, 2008: 199-210

[16] Reidt S, Srivatsa M, Balfe S, et al. The fable of the bees: Incentivizing robust revocation decision making in ad hoc networks [C] //Proc of the 16th ACM Conf on Computer and Communications Security. New York: ACM, 2009: 291-302

[17] Alpcan T, Buchegger S. Security games for vehicular networks [J]. IEEE Trans on Mobile Computing, 2011, 10 (2): 280-290

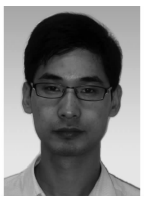
[18] Freudiger J, Manshaei M, Hubaux J, et al. On non-cooperative location privacy: A game-theoretic analysis [C] // Proc of the 16th ACM Conf on Computer and Communications Security. New York: ACM, 2009: 324-327

[19] Fudenberg D, Tirole J. Game Theory [M]. Cambridge: MIT Press, 1991: 207-314

[20] Harsanyi J. Games with incomplete information played by Bayesian players [J]. Management Science, 1968, 14(7): 486-502

[21] Huang Dijiang, Misra S, Xue Guoliang, et al. PACP: An efficient pseudonymous authentication based conditional privacy protocol for VANETs [J]. IEEE Trans on Intelligent Transportations, 2011, 12(3): 736-746

[22] Lin Xiaodong, Sun Xiaoting, Ho P, et al. GSIS: A secure and privacy preserving protocol for vehicular communications [J]. IEEE Trans on Vehicular Technology, 2007, 56(6): 3442-3456

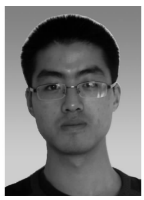


He Yunhua, born in 1987. PhD candidate and student member of China Computer Federation. His current research interests include location and data privacy, security in vehicular ad hoc networks.

Sun Limin, born in 1966. PhD, professor and PhD supervisor. Member of IEEE and senior member of China Computer Federation. His current research interests include wireless sensor networks, vehicular ad hoc network.

Yang Weidong, born in 1977. PhD, associate professor. Senior member of China Computer Federation. His current research interests include wireless networks, information security.

Li Zhi, born in 1985. PhD candidate and student member of China Computer Federation. His current research interests include delay tolerant networks, wireless sensor networks and mobile computing.



Li Hong, born in 1989. PhD candidate and student member of China Computer Federation. His current research interests include location privacy.