基于联合熵隐私保护的自适应动态 Mix-zone 方案

冯霞[1,2]，刘亚伟[3]

（1. 江苏大学汽车与交通工程学院， 江苏 镇江 212013；
2. 江苏省工业网络安全技术重点实验室，江苏 镇江 212013；
3. 安徽大学计算机科学与技术学院， 安徽 合肥 230601）

摘 要：针对车联网中 Mix-zone 方案灵活性低以及隐私保护程度对用户缺乏透明度的问题，提出一种交通自适应的动态 Mix-zone 创建方法，可以根据道路交通状况为车辆动态创建 Mix-zone，随时随地为车辆创建 Mix-zone 进行假名更换，建立基于身份和位置的隐私保护；提出对 Mix-zone 进行隐私分级的联合熵度量模型，可通过归一化的定量计算结果度量 Mix-zone 达到当前区域车辆隐私需求的程度。使用深圳市某区的出租车辆的轨迹数据验证了联合熵隐私度量模型及基于该模型的 Mix-zone 创建方案，实验结果表明，该联合熵模型能刻画交通场景中参数与隐私保护程度的正比关系，在联合熵所表示的无序性指标上，所提 Mix-zone 创建方案相较其他方案，具有更好的隐私保护效果。

关键词：车联网；Mix-zone；联合熵；隐私保护

中图分类号：TP391

文献标识码：A

doi: 10.11959/j.issn.1000-436x.2018038

Dynamic Mix-zone scheme with joint-entropy based metric for privacy-perserving in IoV

FENG Xia[1,2], LIU Yawei[3]

1. School of Automotive and Traffic Engineering, University of Jiangsu, Zhenjiang 212013,China
2. Jiangsu Key Laboratory of Security Technology for Industrial Cyberspace, Zhenjiang 212013, China
3. School of Computer Science and Technology, University of Anhui, Hefei 230601, China

Abstract: Aiming at the weak flexibility and lack of users' transparency existing in the current Mix-zone schemes for Internet of vehicle (IoV), a dynamic was proposed for Mix-zone construction with traffic adaption, which could construct a Mix-zone for the vehicles dynamically according to the traffic conditions for changing pseudonym at anytime and anywhere. This kind of Mix-zone could achieve privacy-preserving based on the identity and location. In addition, a novel traffic-adaptive metric was presented for classifying the privacy leveled in Mix-zone, which applied the normalization quantitation to measure the degree of Mix-zone's privacy demanding by the current region. It was verified that the joint entropy-based privacy measuring model and the Mix-zone construction scheme by utilizing the trajectory data of taxis in certain district in Shenzhen city. The experimental shows that the proposed combination entropy-based model could depict the proportional relationship between the traffic scene parameters and the privacy-preserving degrees. The scheme is better in performance over the related methods, and strikes a good balance between location privacy and service usability.

Key words: Internet of vehicle, Mix-zone, joint-entropy, privacy-preserving

---

收稿日期：2017-09-30；**修回日期**：2018-01-20

基金项目：国家自然科学基金资助项目（No.U1736216, No.61472001, No.61702233）；江苏省重点研发计划基金资助项目（No.BE2015136）

Foundation Items: The National Natural Science Foundation of China (No.U1736216, No.61472001, No.61702233), The Key Research and Development Plan Project of Jiangsu Province (No.BE2015136)

1　引言

车联网（Internet of vehicle）融合现代通信与网络技术，通过人、车、路、云平台实时关联与感知，实现智能交通系统。比较典型的有基于位置的服务（LBS, location-based service）、驾驶辅助和事故警告等。在车联网中，车辆周期性地广播交通车辆的当前位置、车辆速度、车流情况等相关信息给其周围的所有车辆，恶意车辆可以通过关联消息与发送者，获取车辆驾驶者的隐私信息，这对车辆用户的隐私造成潜在的威胁[1~4]，极大阻碍了车联网的推广应用。

针对车联网中的用户隐私保护技术，主要有群签名、Mix-zone（淆乱区域）这 2 类实施方案，这些方案的本质是利用密码学或映射关系来淆乱车辆真实 ID、位置以及用户名之间的关联。签名方案依靠公钥加密技术实现了身份认证和轨迹隐私保护统一，但基于密码学匿名认证技术挑战了车辆网络认证的物理瓶颈，即路侧单元（RSU, road side unit）作为车联网服务的接入点计算和通信能力有限。文献[5]设计了一种批量认证的密钥管理机制，将认证效率提高到 250 次/秒，文献[6]更是采用代理认证的新模式将 RSU 签发认证证书的速度提高到 6 200 次/秒，但是频繁地认证和用户的隐私保护需求依然在挑战 RSU 物理能力的瓶颈。Beresford[7]在 2004 年提出淆乱区域（Mix-zone）的概念，Buttyan 等[8]首次将 Mix-zone 方法用于车联网，其通过建立特定淆乱区域的方法实现了隐私保护。Ying 等[9]允许车辆根据需求提出假名更换需求，系统通过创建 Mix-zone 完成该请求。随后，淆乱区域成为车辆隐私保护的热门技术，其构建的方式也涌现出诸如基于特殊位置[10]、安静时段[11]、加密空间[12,13]和基于通信代理[13]等几种典型方式。这些方案都重点关注隐私保护的效果，不同程度地忽略了 Mix-zone 的创建效率的考虑，在 Mix-zone 的创建效率和隐私保护效果的矛盾选择中缺乏突破性的方法。

本文提出一种自适应的动态 Mix-zone 创建方法，以基于联合熵的安全隐私模型对 Mix-zone 进行隐私度量，而车辆可以根据道路交通状况判断 Mix-zone 的保护等级，进而决定是否进行动态的 Mix-zone，通过对 Mix-zone 分级和用户需求的个性化差异，实现了创建效率和隐私需求的矛盾统一。

2　相关工作

车联网中，隐私保护需求主要体现在通信过程中，需要隐藏通信双方的身份、位置、信息内容等来淆乱车辆身份和位置关联关系。常用的技术有基于签名的方案和基于 Mix-zone 的假名方案，前者用密码学的签名方式淆乱用户的真实 ID，从身份匿名来保护隐私；后者通过淆乱车辆与标记的对应关系，来实现隐私保护。假名方案的效率主要来自签名认证的模式和算法，属于纯密码学的内容。本文主要考虑在 Mix-zone 方案中折中创建效率与隐私保护效果。当前主要的 Mix-zone 方案包括以下 6 类。

1) 基于特殊位置的 Mix-zone，即在事先指定的某个特殊地理区域内更新假名，以达到车辆混淆和隐藏目的。Lu 等[10]提出在车辆聚集地区，如交通信号等处、大型停车场等社交点建立 Mix-zone，同时用博弈论完成证明。但是，因为只能在固定地点更换假名，Mix-zone 的创建不够灵活，为了满足假名更换需求，甚至有些车辆通过更改行驶路径达到更改假名的目的。Mix-zone 的 K 匿名要求也是固定位置的 Mix-zone 的一个挑战。

2) 基于静默时段的 Mix-zone，即车辆自己随机选择一段时间不发送分组（即静默），静默之后再更新假名。Buttyan 等[11]提出了在车辆低速行驶时设置静默时段的 SLOW 方案。在方案中，SLOW 设置静默时段的位置是红绿灯处。该方案中车辆更换其假名后将进入静默周期，一些对时间敏感的服务有可能无法执行。

3) 加密 Mix-zone，是指通过加密区域的方式来保证更换假名的不可关联性。Dahl 等[12]则在十字路口构造加密网络模型。但方案没有考虑到道路网络对 Mix-zone 创建的影响，假设某 Mix-zone 被道路网络分割成相互独立的部分，则小型 Mix-zone 将无法达到 K-匿名所需安全要求。Guo 等[13]提出通过在选定的路口构建加密 Mix-zone，并由 RSU 负责向进入该区域的车辆分发会话密钥。

4) Mix-zone 通信代理，是指用代理方式来实现在 Mix-zone 内的通信。Scheuer 等[14]针对城市交叉口及高速分叉路口，通过代理来实现在 Mix-zone 内更换假名，从而降低了车辆和 RSU 的开销，以代理的绝对可信假设提升了隐私保护的水平。

5) MobiMix 方案[15,16]，在考虑 Mix-zone 基础

理论的同时，将地理信息、用户的密度和行为统计数据、空间和时间的运动模式都加入了 Mix-zone 模型，提高了攻击检测能力。

6) 混合方案，该类方案综合了 Mix-zone 的假名方案和群签名环签名方案的优势以克服其不足，获得更优的综合性能。文献[17]方案中允许车辆根据需要提出假名更换需求，从而要求系统为其创建 Mix-zone。Emara[18]提出的混合方案，综合应用了 3 种方式，使用假名、Mix-zone 和数字签名的思想。通过组建车辆的群导航提供车辆匿名机制，随机安静时段提高了位置隐私，签名密钥管理实现安全与隐私的平衡，极大减少车辆被位置跟踪的可能性。

本文的基础方案是在 Mix-zone 存续期间引入群签名机制，保证该阶段的通信隐私安全，以避免被攻击者通过收集、分析等手段进行身份的关联。然而本文的主要贡献是提出基于联合熵的 Mix-zone 隐私度量模型，通过对交通场景的分析，提前评估该地区建立 Mix-zone 的隐私保护效果，为系统构建最优 Mix-zone 最大限度保护车辆隐私提供判断依据。

3 网络结构与形式化描述

本文采用文献[19]给出的基于位置服务的系统模型，并在此基础上，给出基于无向带权图的形式定义，为后续对 Mix-zone 进行基于联合熵的隐私度量提供基础。

3.1 车联网体系结构

车联网主要包括 4 个部分：可信机构（TA, trusted authority）、基于位置服务（LBS, location based service）、RSU 以及装配有车载单元（OBU, on board unit）的车辆（用 V 来表示）。如图 1 所示，各部分基本功能如下。

1) 可信机构（TA）

服务器 TA 是绝对安全的信任实体，负责生成系统全局的安全参数、为各参与方颁发公私钥以及车辆的认证参数，TA 存储着所有经过认证的车辆用户的身份信息，经过 TA 授权后，可以调查车辆的既往位置信息等，因此，在本文方法中需要协助车辆构建动态 Mix-zone，它提供认证、签名和集中式的管理服务。

2) 路侧单元（RSU）

RSU 一般通过无线网络与车辆进行通信，同时通过有线方式与其他 RSU、TA 通信，RSU 网络是整个车联网的骨干通信网，车辆是通过 RSU 接入网络服务的。RSU 架设在通行道路的主要路口，每个 RSU 都能接收车辆的注册请求信息，对通过验证的车辆提供服务。本文假设一共有 m 个 RSU，记为

$$\boldsymbol{R}=\{R_1,R_2,\cdots,R_m\} \tag{1}$$

3) 车辆（V）

每个车辆上装有一个 OBU 模块，通过无线方式和其他车辆或 RSU 通信，获得网络系统的服务；发送已知交通信息给系统。本文所述系统中，车辆真实身份只有 TA 与车辆自身知道。假设网络中有 n 辆车，车辆集合形式化表示为

$$V=\{V_i \mid 1\leqslant i\leqslant n\} \tag{2}$$

4) 基于位置服务（LBS）

LBS 是车联网中的位置应用服务器，车辆将位置服务请求发送给 LBS，服务器对车辆的请求内容分析处理过后，把位置服务数据通过 RSU 反馈给车辆。

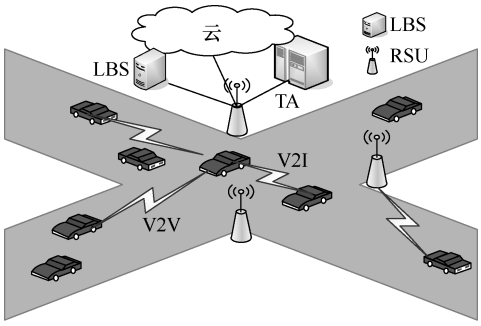


图 1 车联网结构

3.2 无向带权图模型

交通场景的某一个区域，若准备设置成 Mix-zone，为了表述清晰，可以描述成以 RSU 为顶点的无向带权连通图，具体定义如下。

定义 1 交通图中的任意一个区域可以描述成具有 n 个顶点的无向带权连通图 ***Zone***

$$\boldsymbol{Zone}=(Vertex,Edge,W,d) \tag{3}$$

其中，d 为图 ***Zone*** 的点连通度，顶点集合 $Vertex$ 是由式 (1) 所定义集合 RSU 的子集，其中 $|Vertex|=n, n\ll m$；对 $Vertex$ 中任一顶点 R_i 的点权记为 $|R_i|$，$Edge$ 是图 ***Zone*** 中连接不同顶点 RSU 的边集合，$Edge$ 中每一个成员 e 可以表示为 $Vertex$ 中顶点组成的二元组。

$$Edge=(R_i,R_j) \tag{4}$$

其中，$R_i,R_j\in Vertex$。

W 是与 $Edge$ 对应的图 M 中边的权值集合，式(4)所定义边 $Edge$ 的权值 $|Edge|$ 记为

$$|Edge|=W(R_i,R_j) \tag{5}$$

在不引起混淆的情况下，$W(R_i,R_j)$ 通常简写为 $W(i,j)$ 或 w_{ij}。

在车联网交通环境里，本文用定义 1 所给出的点权表示该 RSU 通信范围内车辆的数目，用边权表示该道路上所能正常通行的最多车辆数目，即在行车经验上一个理论的最大值。参考文献[20]对城市车辆移动轨迹的统计和分析结论，车辆达到道路的状况遵循泊松概率分布过程，则边权就是该泊松分布的顶点值。显然，交通无向带权图具有性质 1 的特点。

性质 1　在定义 1 所述的无向带权连通图 ***Zone*** 中，如果 2 个顶点 R_i、R_j 是没有直接连通的路，则该边权 w_{ij} 的值为 0。

为了定义具备数学意义上的完备性，本文定义零边 (R_i,R_j)，当 $i=j$ 时，边权 w_{ii} 为

$$w_{ii}=\sum_{k=1,k\neq i}^{n}w_{ik} \tag{6}$$

其中，w_{ii} 表示通过 R_i 的最大车辆数目，而计算中其实际数值有可能超过 RSU 节点 R_i 通信范围内所能容纳的最大车辆数据，因此，式(6)仅具有数学计算一致性的理论价值，不能和实践中具体情况对应。为了更好地理解定义 1，本文以图 2 为例给出了一个具体的示例，图 2(a)中所示的交通环境，参照定义 1 转化为图 2(b)所示的无向带权图，其边和点的权值取决于瞬时道路上和 RSU 覆盖范围内车辆的数目。如图 2(b)中节点 5 覆盖范围内的车辆有 2 个，即权值为 2，在图 2 中表示为 2)，而节点 1 和节点 4 的连接边的边权 w_{14}，虽然当前只有一辆车，但是其边权需要根据既往交通记录及道路状况设定。

3.3　Mix-zone 的形式化描述

Mix-zone 是一个典型的交通区域，符合定义 1 所述的形式化记录。然而，Mix-zone 作为淆乱车辆身份的功能区，主要体现在交叉口，因为在道路上，车辆在时序上的规律性很容易被跟踪。因此，Mix-zone 的隐私保护能力主要由 3 个值确定，分别是该区域所能容纳的最大平均交通流、每个顶点停留的瞬时车辆数目以及车辆在该区域各顶点的平均停留时间。

如果将定义 1 所述的无向带权图作为 Mix-zone，则其所能容纳的最大平均交通流 $D_{\max}$ 可以表示为

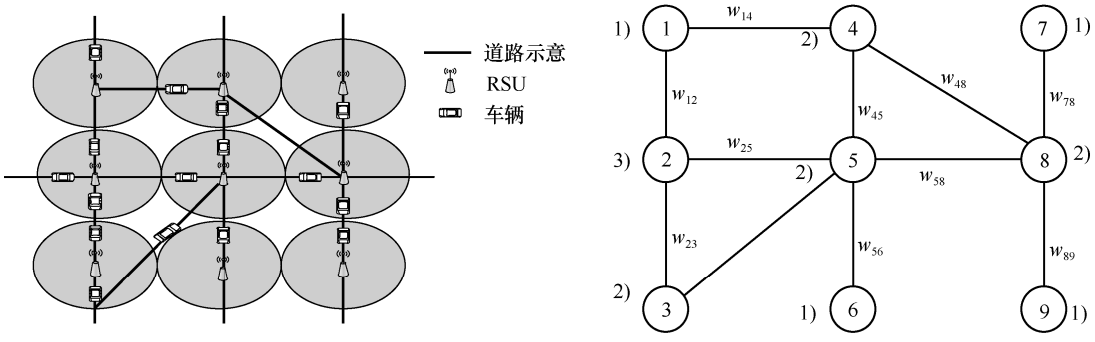
$$D_{\max}=\frac{\sum_{i=1}^{n}w_{ii}}{d\sum_{i,j=1,i\neq j}^{n}|w_{ij}|} \tag{7}$$

其中，w_{ij}、w_{ii} 分别由式(5)和式(6)给出，而 $|w_{ij}|$ 定义为

$$|w_{ij}|=\begin{cases}1, w_{ij}\neq 0\\ 0, w_{ij}=0\end{cases} \tag{8}$$

对某一特定 Mix-zone，$D_{\max}$ 与全区域对应的交叉口平均通过车辆数目成正比关系，因此，其值越大，则该 Mix-zone 所能容纳的最大平均交通流越大、所能保护车辆与假名对应隐私的能力越强。

文献[20]通过对城市出租车运动轨迹的分析，



(a) 交通场景实例　　(b) 无向带权图 ***Zone***

图 2　交通区域与形式化权图示例

获得车辆到达某一个汇聚点的概率遵循泊松分布。由此可以假定车辆到达某一 RSU 的间隔时间 t_a 服从参数为 $\frac{1}{\lambda}$ 的指数分布。令 X 为 t 时刻 Mix-zone 的车辆总数目。因此，概率 $X=x$ 在 t_a 期间可以表示为

$$P_r[X=x|t=t_a]=\frac{(\lambda t)^x}{x!}\mathrm{e}^{(-\lambda t)} \tag{9}$$

攻击者根据车辆旧的假名（驶进 Mix-zone 之前）映射到新的假名（驶出 Mix-zone 之后），如果 t 时刻是完全无依据猜测，则每辆车被猜中的概率为 $\frac{1}{x}$。假设攻击者获得了某些或对或错的信息，对某车辆 V_i 旧假名对应驶出 Mix-zone 车辆进行有区别的概率赋值，将该假名对应到车辆 V_j 的概率为 $\boldsymbol{M}_{ij}$，则有

$$\sum_{j=1}^{x}\boldsymbol{M}_{ij}=1 \tag{10}$$

其中，x 是 Mix-zone 中 t 时刻的车辆数。$\boldsymbol{M}_{ij}$ 是攻击者在 Mix-zone 系统成功猜测出车辆 V_i 真实身份的概率值，该值与 $\frac{1}{x}$ 的接近度和 x 的大小、该时刻车辆数目为 x 的概率 $P_r\left[X=x|t=t_a\right]$ 共同决定了该时刻 Mix-zone 的车辆隐私保护能力的大小。

由于 t 时刻 Mix-zone 的车辆总数是 x，则每辆车的存续时间以向量形式表示为

$$\boldsymbol{\Gamma}_v=[\tau_1,\tau_2,\cdots,\tau_x]$$

由于式(9)已经给出了该时刻车辆数目为 x 的概率值，因此，如果假设 Mix-zone 中持续存在车辆数据低于 x，则 Mix-zone 结束服务，那么一个 Mix-zone 的存续时间 Δt 可以定义为车辆存在时间的平均值，而非最小值。

$$\Delta t=\sum_{i=1}^{x}\frac{\tau}{x} \tag{11}$$

3.4 Mix-zone 的隐私特征与车辆隐私期望

Mix-zone 中车辆匿名的主要需求是 K 匿名，按照式(9)及式(11)的定义，Mix-zone 创建方案就是要保证在其区域范围内从概率意义上至少含有 k 辆车 $(x>k)$，而且要从最低要求上保证在 Δt 时间内最初的 k 辆车不会离开。因为根据 Mix-zone 随时创建的特性，参与创建的 k 辆车中任何一辆离开 Mix-zone，该 Mix-zone 将解散。虽然在 Mix-zone 的存续期间，允许其他车辆随机进入或离开 Mix-zone，但是这些车辆面临攻击者蓄意的分析，将很容易被排除，因而最初存在的 k 辆车将是攻击者通过时间分析后仍然具有隐私保护价值的最小子集。

设 Mix-zone 中车辆集合为 V，而不同车辆对隐私等级有不同需求，集合 V 中车辆对应的隐私等级需求为

$$\boldsymbol{C}=\left\{c_1,c_2,\cdots,c_{|V|}\right\} \tag{12}$$

其中，$|V|$ 是 Mix-zone 中车辆的总数目，而 c_i 对应车辆 V_i 的隐私需求值。c_i 取值与 K 匿名中的 $\frac{1}{k}$ 值关联。本文实验把车辆的匿名需求分为 3 个等级：零需求、普通需求和高需求。基本定义为

$$c(V_i)=\begin{cases}\frac{1}{2}<c_i\leqslant 1, & \text{零需求}\\ \frac{1}{k}<c_i\leqslant\frac{1}{2}, & \text{普通需求}\\ 0<c_i\leqslant\frac{1}{k}, & \text{高需求}\end{cases}$$

即部分车辆对隐私保护无需求，一些只要有 $\frac{1}{2}$ 的混淆概率即可，另外一些车辆则要求更高，其值为 $\frac{1}{k},k>>2$。本文实验所对应的 3 个等级，分别取值为 1、$\frac{1}{2}$、$\frac{1}{4}$，则车辆 i 的隐私需求值 c_i 为 $\frac{1}{4}$、$\frac{1}{2}$、1 中某一个数值。

另外，3.3 节已经把区域 Mix-zone 的隐私保护能力分为 3 个值：所能容纳的最大平均交通流、每个顶点停留的瞬时车辆数目以及车辆在该区域各顶点的平均停留时间，并对应符号化为 $D_{\max}$、X 和 Δt 这 3 个数量指标，不妨将该指标对应为以下三元组。

$$\boldsymbol{\Delta}=\{\delta_1,\delta_2,\delta_3\}$$

一般地，记 $\boldsymbol{\Delta}$ 为 m 元组，$\boldsymbol{\Delta}=\{\delta_1,\delta_2,\delta_3\}$，此时，若记 Mix-zone 中车辆数目 $|V|=n$，则 Mix-zone 隐私等级评价特征值矩阵为

$$\boldsymbol{Y}_{m\times n}=\begin{bmatrix}y_{11} & y_{12} & \cdots & y_{1n}\\ y_{21} & y_{22} & \cdots & y_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ y_{m1} & y_{m2} & \cdots & y_{mn}\end{bmatrix} \tag{13}$$

其中，y_{ij} 表示车辆节点 V_j 指标的特征值。

至此，获得了 2 个矩阵、一个向量，分别是式(10)表示的车辆 V_i 旧假名对应驶出 Mix-zone 的车辆 V_j 的概率矩阵 $\boldsymbol{M}_{ij}$、式(13)表示的 Mix-zone 隐私等级评价特征值矩阵和式(12)表示的 Mix-zone 中车辆的隐私等级需求向量。在计算逻辑上，式(13)决定了式(10)的猜测概率，而式(10)中矩阵 $\boldsymbol{M}$ 的对角线应高于式(12)中向量 $\boldsymbol{C}$ 的对应元素值。即

$$\boldsymbol{M} = f(\boldsymbol{Y}) \tag{14}$$

约束条件为 $\sum_{j=1}^{n}\boldsymbol{M}_{ij}=1$，$\forall i$，$i\in\{1,\cdots,n\}$。

而该时刻 Mix-zone 满足用户需求的判断式为

$$\boldsymbol{M}_{ij} \geqslant \boldsymbol{c}_i, \forall i, i\in\{1,\cdots,n\} \tag{15}$$

由于隐私需求向量中的值被等级化，且与 k 值唯一关联，因此，$f(*)$ 函数的目标是把 $\boldsymbol{Y}$ 矩阵转化为等级相关的数值内容，以联合熵来定义 $f(*)$ 函数并度量 Mix-zone 的隐私保护能力。更简单的处理方式是把式(14)和式(15)所定义的判断过程，转化为每辆车辆采用式(16)的形式，判断每个 Mix-zone 是否符合自身的隐私保护需求。

$$\min\left(\left(\sum_{\substack{x=k\\ t_a\leqslant\Delta t}}^{n} P_r\left[X=x|t=t_a\right]\right), D_{\max}\right) \geqslant c_i \tag{16}$$

由于 c_i 的赋值可以是车辆隐私保护的等级，因而，式(16)可用于判断 Mix-zone 对车辆隐私保护能力进行标准化的等级判定。第 4 节用联合熵来描述判断过程中的不确定性，并基于对相应时刻 K 值与每个车辆对所属等级概率判断，来对整个 Mix-zone 隐私保护等级能力进行度量。表 1 给出了主要参数及其定义。

表 1　主要参数

标识符	描述
$P_r[X\|t]$	t_a 时刻车辆进入 Mix-zone 的概率
$\boldsymbol{\Gamma}_v$	车辆的存续时间
$\boldsymbol{C}$	Mix-zone 的等级域
$\boldsymbol{\Delta}$	车辆隐私保护影响因子向量
$\boldsymbol{M}_{ij}$	车辆节点隐私评估概率矩阵
$\boldsymbol{A}_{l\times n}$	Mix-zone 指标标准化矩阵
Δt	Mix-zone 存续时长
H	Mix-zone 联合熵
$C(\text{Mix-zone})$	Mix-zone 隐私等级

4　基于联合熵的 Mix-zone 隐私模型

式(13)给出了该 Mix-zone 内部节点及其对 n 个影响因素的评估阵列，进一步将系统的评估值等级化，记为 l 个等级，例如，分为 0、$\frac{1}{2}$、1 这 3 个等级，则 $l=3$。可将 m 个因素的值对应标准化为 l 个不同数值，对同等级数值进行聚类，并将各个因素对同一节点的影响权重归一化，即将对同一节点的各影响因素按照其影响程度定义百分比，从而节点 j ($(1\leqslant j\leqslant n)$)以 α_{ij} 概率属于 h ($(1\leqslant h\leqslant l)$)级 Mix-zone 隐私等级标准。

$$\boldsymbol{A}_{l\times n}=\begin{bmatrix}\alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n}\\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ \alpha_{l1} & \alpha_{l2} & \cdots & \alpha_{ln}\end{bmatrix} \tag{17}$$

其矩阵元素 α_{hj} 满足式(18)所定义的约束条件。

$$\sum_{h=1}^{l}\alpha_{hj}=1, \forall j, 0\leqslant\alpha_{hj}\leqslant 1 \tag{18}$$

这样，整个 Mix-zone 系统的联合熵定义为

$$\boldsymbol{H}=-\sum_{j=1}^{n}\sum_{h=1}^{l}[\alpha_{hj}\ln(\alpha_{hj})] \tag{19}$$

整个 Mix-zone 系统的隐私保护等级为

$$C(\text{Mix-zone})=\frac{H}{n\ln(l)} \tag{20}$$

同样，对影响车辆节点隐私保护内容的指标向量 $\boldsymbol{\Delta}$ 进行归一化，使其满足各因素综合影响力为 1，如式(21)所示。

$$\boldsymbol{\Delta}=\delta_1,\delta_2,\cdots,\delta_m \tag{21}$$

其中，$\sum_{i=1}^{m}\delta_i=1$。

为了确定合理的 α_{h_j}，采用汉明距离定义车辆节点 j 与 Mix-zone 隐私保护等级 h 的差异，为

$$e_{hj}=\sum_{i=1}^{m}\left((\delta_i\times\frac{1}{n}\times\sum_{p=1}^{l}\left|m_{jj}-m_{pj}\right|\right) \tag{22}$$

Mix-zone 隐私保护等级保护评价系统的目的是使用户根据交通场景，选择新创建的 Mix-zone 的时机，使隐私保护等级与标准定义之间的加权广义距离之和最小，即

$$\min_{\Delta} E = -\sum_{j=1}^{n}\sum_{h=1}^{l}\alpha_{hj} \times e_{hj} \tag{23}$$

其中，$\sum_{i=1}^{m}\alpha_{hj}=1, \forall j, 0 \leqslant \alpha_{hj} \leqslant 1$。

为了提高匿名需求中的模糊性和不确定性，根据 Jaynes 最大熵[21]原理，应使 $\min_{\alpha_{hj}} e$ 最大，即

$$\min_{\alpha_{hj}} e = -\sum_{j=1}^{n}\sum_{h=1}^{l}[-\alpha_{hj}\ln(\alpha_{hj})][\tag{24}$$

其中，$\sum_{i=1}^{m}\alpha_{hj}=1, \forall j, 0 \leqslant \alpha_{hj} \leqslant 1$。

因此，求最优 Mix-zone 是一个双目标优化问题，为解决此问题，构造双目标函数，最终获得本文的熵极大化隐私保护模型。

$$H = \min_{\alpha_{hj}} \sum_{j=1}^{n}\sum_{h=1}^{l}\alpha_{hj}e_{hj} + \frac{1}{K}\sum_{j=1}^{n}\sum_{h=1}^{l}[-\alpha_{hj}\ln(\alpha_{hj})] \tag{25}$$

其中，$\sum_{i=1}^{m}\alpha_{hj}=1, \forall j, 0 \leqslant \alpha_{hj} \leqslant 1, j=1,2,\cdots,n$，$K$ 是目标系统所能提供的最大 K 匿名的匿名值。

5 自适应动态 Mix-zone 创建方案

本节首先描述了基础的 Mix-zone 创建算法，该算法可获得 Mix-zone 度量的初始输入值。然而，为了获得最优的 Mix-zone 性能，需要创建方法支持 Mix-zone 持续期间，车辆节点之间使用群签名认证进行适度通信，而不影响假名更新的隐私保护性能；然后，在基础算法中嵌入群签名协议，可以支持车辆自适应动态创建 Mix-zone 的选择。

5.1 基础方案概述

假设车辆 V_i 需要更换假名，设其在某一 RSU 范围内。车辆 V_i 发送一个假名更换请求，TA 收到假名更换请求后将为该车辆执行创建算法。本文提出一种自适应 Mix-zone 创建方案。首先，令 V_{MZ} 为一个空集合，将车辆所在的 RSU 顶点加入集合；其次，遍历顶点集合的相邻顶点 V_{AZ}，同步将车辆 V_i 的假名使用时间 t 记录下来；最后，提取最长假名时间顶点 AZ_t，合并到顶点集合，生成新的 V_{MZ}，达到匿名要求时算法结束。算法获得的顶点集合就是最新的 Mix-zone。为了防止车辆频繁更换假名，所以在方案中考虑了假名使用时间。算法 1 描述了这种启发式方案的动态创建过程。获得 Mix-zone 后，本文分别计算出 $D_{\max}$、$P_r[X|t]$ 的值，用于下一步度量该 Mix-zone 的隐私等级。

算法 1 动态 Mix-zone 创建

输入 无向带权图 $\boldsymbol{Zone}=(Vertex, Edge, W, d)$, $\boldsymbol{R}$

输出 Mix-zone

1) RSU 集合 $V_{MZ} \leftarrow \varnothing$；
2) 遍历顶点 V_{AZ} 的所有相邻节点；
3) 初始化车辆数目 $V_{MZ}, N \leftarrow 0$；
4) 当 $N \leqslant k$ 时
5) 对于所有的 $AZ_t \in V_{AZ}$
6) 计算车辆的假名使用时间 $AZ_t : T_t = (t_1 + t_2 + \cdots + t_n)$

其中，n 是在 AZ_t 中的车辆数目

7) 结束
8) 选择拥有最长 T 的车辆；
9) $V_{MZ} \leftarrow V_{MZ} \cup AZ_t$；
10) $N \leftarrow N + n$；
11) 对于所有的 $MZ_t \in List\langle V_{MZ}\rangle$
12) 通过式（7）计算 $D_{\max}$；
13) 通过式（9）计算 $P_r[X|t]$；
14) 结束
15) 结束

算法 1 满足了动态创建 Mix-zone 的要求，但其不能保证生成的 Mix-zone 最大化车辆的隐私保护机制。5.2 节将根据基础方案提出新型的基于联合熵的隐私度量方法，并根据该度量方法实现最优动态 Mix-zone 的创建。

5.2 基于联合熵的动态 Mix-zone 创建方案

为了最大化 Mix-zone 的联合熵，即创建最优 Mix-zone。本文根据当前道路中交通状况进行 Mix-zone 创建的同时，必须实时计算出其联合熵，动态选择最优 Mix-zone。基于联合熵隐私保护的 Mix-zone 方案主要由 3 个阶段组成：车辆证书初始化、选择最优 Mix-zone 和基于群签名的 Mix-zone 期间的通信。

1) 车辆证书初始化。如图 3 所示，采用 Dan-Byen 短签名[22]方案用于系统初始化。车辆加入车联网系统时首先需要向 TA 注册，注册过程采用文献[12]的方案。车辆 A 计算得到自己的公钥 PK_A 和假名 $Pesu\_IDA$，并将它们发送给所在区域的路边单元 R；本地 RSU 收到 A 发送的消息后，向 TA 转发车辆 A 的公钥和假名，查询 A 的公钥和

假名的正确性；TA 根据所持有的与车辆 A 共有的主密钥 s 验证车辆 A 的假名，如果正确，给路边单元 R 回复一个确认消息；RSU 接收到 TA 的确认消息后，计算出车辆 A 的证书 $Cert(R,v)$ 存在自己的证书列表（CL, certificate list）里，同时发送出自己的 RSU 证书给 A，作为车辆 A 生成其临时证书 $Cert(R,v)$ 的资料。

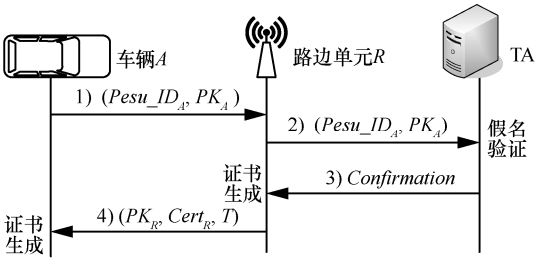


图 3　车辆证书初始化

2) 择选最优 Mix-zone。车辆 V_t 向系统申请更换假名，V_t 向 TA 发出更换假名的申请，TA 接收申请后将为车辆 V_t 执行 Mix-zone 创建程序。根据 4.1 节的分析可知，TA 在执行创建 Mix-zone 的算法过程中不仅要保证 K 匿名，还要求 Mix-zone 的联合熵最大化。因此，根据 5.1 节的基础方案需要进一步地考虑安全隐私保护的最大化，选择联合熵最大的 Mix-zone。构建交通自适应的动态 Mix-zone 过程如算法 2 所示。

算法 2　基于联合熵的动态自适应 Mix-zone 创建

输入　无向带权图 $\boldsymbol{Zone}=(Vertex,Edge,W,d)$, $\boldsymbol{R}$

输出　Mix-zone

① 拥有 k 辆车的 Mix-zone= $List\langle V_{MZ}\rangle$；

② 深度优先搜索

③ Mix-zone RSU 集合 $V_{MZ}\leftarrow\varnothing$；

④ 搜索 V_{AZ} 集合的相邻节点；

⑤ 初始化车辆集合 $V_{MZ}, N\leftarrow 0$；

⑥ 当 $N\leqslant k$ 时

⑦ $V_{MZ}\leftarrow V_{MZ}\cup AZ_t$；

⑧ $N\leftarrow N+n$；

⑨ 结束

⑩ 将 V_{MZ} 合并到 $List\langle V_{MZ}\rangle$；

⑪ 结束

⑫ 对于所有的 $MZ_t\in List\langle V_{MZ}\rangle$

⑬ 通过式(25)计算 H；

⑭ 选择联合熵最大的 Mix-zone；

⑮ 结束

3) 基于群签名的 Mix-zone 期间的通信。根据 2.3 节的分析，在 Mix-zone 存续阶段，引入群签名机制。根据本文方案，车辆随机地进入或离开 Mix-zone，RSU 负责分发管理群身份 GID 以及相关的密钥证书。群管理者 GL_j 验证车辆 V_i 的合法性。车辆进入 Mix-zone 之前使用假名 $Pesu\_ID_A$ 发送消息，进入 Mix-zone 后，群管理者 GL_j 向车辆 V_i 提供群身份 GID 以及相关的密钥和证书。在 Mix-zone 存续期间，车辆 V_i 将会使用群身份 GID 来替代自己的假名 $Pesu\_ID_A$ 并与其他车辆通信，这样将有效防止攻击者对车辆的身份进行跟踪及关联，有效保护车辆隐私。引入时间戳机制，保证消息来源的合法性、消息的时效性。车辆在离开 Mix-zone 之前都会使用包含群私钥和证书的群身份进行安全消息的广播，因此，群签名的方案保证车辆在 Mix-zone 存续期间的正常通信的同时，也对车辆的安全隐私进行了保护。

6　性能分析

本文使用深圳市出租车[23]GPS 轨迹数据集评估动态 Mix-zone 的方案性能。本文获取的数据集一天大约有 80 万条出租车 GPS 数据，每辆出租车的 GPS 数据最小上传时间间隔为 15 s。由于深圳市并未部署车联网，在实验中假设 RSU 在道路网络中是全覆盖的，且交叉路口均设有 RSU。本文算法使用 Python 3.5 实现，仿真计算机硬件配置为 Intel i5 的 CPU 和 8 GB 的内存。

为了分析算法本身的性能，提取车流量信息和满足隐私要求的车辆比例作为实验数据。考虑 2 种极端情形，即 7:00~9:00 的高交通流量和 20:00~22:00 的低交通流量情形，将车辆隐私保护等级平均分为 3 档（即每种需求的车辆各占 $\frac{1}{3}$）：无要求、$\frac{1}{2}$ 要求和 $\frac{1}{5}$ 要求，分析不同路口中所能容纳的最大平均交通流 $D_{\max}$ 对隐私保护效果的影响。在图 4 中，以不同路口实际通行的最大车辆数代表最大交通流，而满足隐私保护要求的车辆比例来衡量隐私保护效果。由图 4 可以看出，车辆越多，隐私效果越好，高交通流情形下，同样多的车辆隐私效果反而不好，由此表明车辆在 Mix-zone 中的存续时间越长，效果越好。

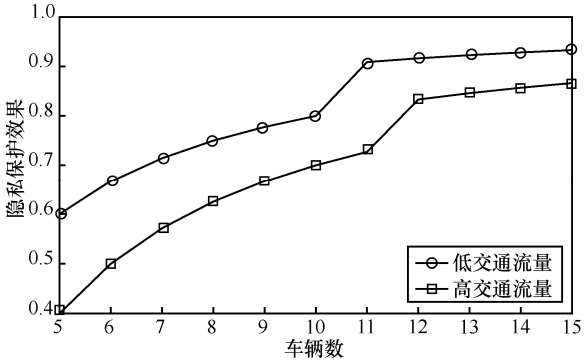


图 4　Mix-zone 交通流对隐私保护效果的影响

图 5 展示了 Mix-zone 存续期间每个顶点停留的瞬时车辆数目和所提供隐私保护性能的关系。同样考虑 7:00~9:00 的高交通流量和 20:00~22:00 的低交通流量这 2 种极端情形，选取不同 Mix-zone，分析每个顶点停留的瞬时车辆数目对满足隐私保护要求的车辆比例的影响。数据表明，瞬时车辆数目越多，车辆存续时间越长，隐私效果越好，这说明本模型描述与真实情况相符。

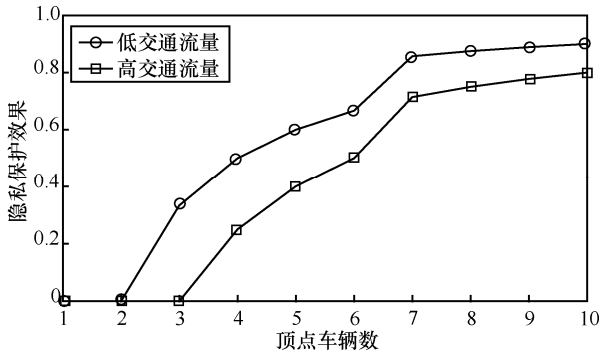


图 5　Mix-zone 权值与隐私保护的关系

图 6 选取了在固定区域，持续时间仅为 7:00~22:00 时，本文方案与文献[5]方案、文献[9]方案等所建立 Mix-zone 的熵值的对比分析，给出了 Mix-zone 区域不同交叉路口停留的瞬时车辆数目、Mix-zone 存续时间与联合熵值之间的关系。需要说明的是，由于本文实验数据来自真实数据，因而“容纳的最大平均交通流”在地图中是固定的，并作为比较因素之一。由于联合熵用来刻画 Mix-zone 区域中的车辆无关联程度：联合熵越大，该区域车辆的身份信息的模糊程度就越高。因此，当该车辆离开 Mix-zone 时，攻击者成功关联车辆新旧假名的概率就越低，从而达到对车辆隐私信息保护的目的。本文系统根据车辆位置及其所处路段的实时交通状况创建 Mix-zone，同时计算联合熵用以预测最佳 Mix-zone，图 6 的显示结果证明了这一点，这说明本文方法体现了理论模型需要追求的效果，且所建立的 Mix-zone 在联合熵值所体现的隐私保护效果上优于其他方法。

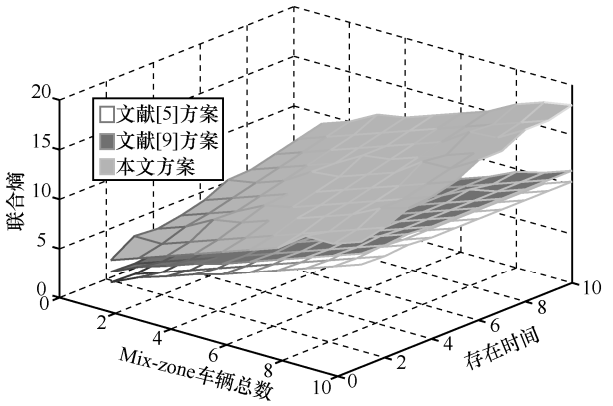


图 6　不同方案中 Mix-zone 车辆总数、存续时间与联合熵之间的关系

7　结束语

本文提出了基于联合熵的自适应动态 Mix-zone 创建方案来保护车辆用户的隐私，旨在解决动态创建的 Mix-zone 中隐私保护效果及用户需求之间的匹配问题，首先，提出基于联合熵的自适应度量方法用于定量评估 Mix-zone 的隐私等级，在动态创建过程中选择联合熵值最优的 Mix-zone 创建方法。以深圳市出租车 GPS 轨迹数据和道路交通图为基础数据来仿真评估所提出方案的性能，结果表明，该评估模型反映了交通因素对隐私效果的实际影响，且所提方法和同类方法相比，在该模型下具有更优的隐私保护效果。

参考文献：

[1] LIU X, ZHAO H, PAN M, et al. Traffic-aware multiple mix zone placement for protecting location privacy[C]//IEEE INFOCOM. 2012: 972-980.

[2] ZHANG L, WU Q, QIN B, et al. Practical secure and privacy-preserving scheme for value-added applications in VANETs[J]. Computer Communications, 2015, 71:50-60.

[3] 刘怡良, 石亚丽, 冯蒿, 等. 车联网中基于神经网络的入侵检测方案[J]. 通信学报, 2014,72(35):233-239.
LIU Y L, SHI Y L,FENG G. Intrusion detection scheme based on neural network in vehicle network[J]. Journal of Communications, 2014, 72(35): 233-239.

[4] FERRAG M, MAGLARAS L, AHMIM A. Privacy-preserving schemes for Ad Hoc social networks: a survey[J]. IEEE Transaction on Communications Surveys & Tutorials, 2017, 19(4):3015-3045.

[5] LU R, LIN X, LUAN T H, et al. Pseudonym changing at social spots:

an effective strategy for location privacy in VANETs[J]. IEEE Transactions on Vehicular Technology, 2012, 61(1):86-96.

[6] LIU Y, WANG L, CHEN H H. Message authentication using proxy vehicles in vehicular ad hoc networks[J]. IEEE Transactions on Vehicular Technology, 2015, 64(8):3697-3710.

[7] BERESFORD A R, STAJANO F. Mix zones: user privacy in location-aware services[C]//IEEE Conference on Pervasive Computing and Communications Workshops. 2004: 127-132.

[8] BUTTYAN L, HOLCZER T, VAJDA I. On the effectiveness of changing pseudonyms to provide location privacy in VANETs[C]// European Conference on Security and Privacy in Ad-Hoc and Sensor Networks. 2007:129-141.

[9] YING B, MAKRAKIS D. Pseudonym changes scheme based on candidate-location-list in vehicular networks[C]//IEEE International Conference on Communications. 2015:7292-7297.

[10] LU R, LIN X, LUAN T H, et al. Pseudonym changing at social spots: an effective strategy for location privacy in VANETs[J]. IEEE Transactions on Vehicular Technology, 2012, 61(1):86-96.

[11] BUTTYAN L, HOLCZER T, WEIMERSKIRCH A, et al. SLOW: a practical pseudonym changing scheme for location privacy in VANETs[C]// IEEE Vehicular Networking Conference. 2010:1-8.

[12] DAHL M, DELAUNE S, STEEL G. Formal analysis of privacy for vehicular mix-zones[C]//European Symposium on Research in Computer Security Computer Security(ESORICS2010). 2010: 55-70.

[13] GUO N, MA L, GAO T. Independent mix zone for location privacy in vehicular networks[J]. IEEE Access, doi: 10.1109/ACCESS.2018.2800907.

[14] SCHEUER F, FUCHS K P, FEDERRATH H. A safety-preserving mix zone for VANETs[C]//International Conference on Trust, Privacy and Security. 2011:37-48.

[15] PALANISAMY B, LIU L. Attack-resilient mix-zones over road networks: architecture and algorithms[J]. IEEE Transactions on Mobile Computing, 2014, 14(3):495-508.

[16] PALANISAMY B, LIU L. MobiMix: protecting location privacy with mix-zones over road networks[C]//International Conference on Data Engineering. 2011: 494-505.

[17] SAMPIGETHAYA K, LI M Y, HUANG L P, et al. Amoeba: robust location privacy scheme for VANET[J]. IEEE Journal on Selected Areas in Communications, 2007, 25(8): 1569-1589.

[18] EMARA K. Safety-aware location privacy in VANET: evaluation and comparison[J]. IEEE Transactions on Vehicular Technology, 2017, 66(12): 10718-10731.

[19] ZENG S, CHEN Y, TAN S, et al. Concurrently deniable ring authentication and its application to LBS in VANETs[J]. Peer-to-Peer Networking and Applications, 2016, 10(4):1-13.

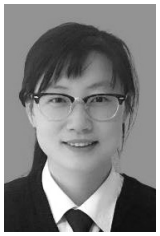
[20] HOSMA M. A study of the source traffic generator using poisson distribution for ABR service[J]. Modelling and Simulation in Engineering. 2012 (1):1-6.

[21] JAYNES E T. Information theory and statistical mechanics[J]. Physical Review,1957,106(4):620-630.

[22] DAN B, BOYEN X, SHACHAM H. Short group signatures[J]. Advances in Cryptology-CRYPTO, 2004, 22(6):41-55.

[23] 刘亚伟. VANETs 中安全认证与隐私保护的研究[D]. 合肥: 安徽大学, 2017.
LIU Y W. Research on security authentication and privacy preservation in VANETs[D]. Hefei: Anhui University,2017.

[作者简介]



冯霞（1983-），女，江苏镇江人，博士，江苏大学讲师，主要研究方向为车联网安全。

刘亚伟（1994-），男，安徽阜阳人，安徽大学硕士生，主要研究方向为车联网安全、云计算等。