·网络空间安全· 文章编号：1000-3428(2021)01-0129-10 文献标志码：A 中图分类号：TN915.08

车联网中基于车辆行为预测的身份认证方案

杨雪婷[1]，李 重[1,2]

（1.东华大学 信息科学与技术学院，上海 201620；2.同济大学 嵌入式系统与服务计算教育部重点实验室，上海 201804）

摘 要：车联网中传统基于密码学的身份认证方案可满足车辆身份认证的基本要求，但其作为静态防御机制不能有效解决车辆身份盗用和认证低时延问题。在基于移动边缘计算框架的软件定义车联网体系结构下，提出一种基于车辆行为预测的身份认证方案。在车辆历史行为数据的基础上，使用前缀树确定认证基站，采用决策树算法和多元非线性回归模型提前对车辆到达站点和时间进行预测，并通过对比车辆到达站点和时间的真实值与预测值实现车辆身份认证。实验结果表明，该方案利用软件定义网络的集中式全局控制能力和移动边缘计算的分布式计算能力对车辆身份认证任务进行管理和分配，可在保证较高车辆认证准确率的同时满足车联网的低时延需求。

关键词：车联网；认证；行为预测；软件定义网络；移动边缘计算

开放科学(资源服务)标志码(OSID)：

中文引用格式：杨雪婷，李重.车联网中基于车辆行为预测的身份认证方案[J].计算机工程，2021，47(1)：129-138.

英文引用格式：YANG Xueting，LI Zhong. Identity authentication scheme based on vehicle behavior prediction for IoV[J].Computer Engineering，2021，47(1)：129-138.

Identity Authentication Scheme Based on Vehicle Behavior Prediction for IoV

YANG Xueting[1]，LI Zhong[1,2]

（1.College of Information Science and Technology，Donghua University，Shanghai 201620，China；
2.Key Laboratory of Embedded System and Service Computing，Ministry of Education，Tongji University，Shanghai 201804，China）

【Abstract】Most of the cryptography-based authentication schemes in Internet of Vehicles（IoV）can meet the basic requirements of vehicle identity authentication，but as the static defense mechanism they cannot effectively solve the problem of identity theft and low latency of authentication. To address the problems，this paper proposes a identity authentication scheme based on vehicle behavior prediction for software defined IoV within the Mobile Edge Computing（MEC）framework.The scheme uses the prefix tree to determine the authentication base station according to the history data of vehicle behavior.Then the decision tree algorithm and the multiple nonlinear regression model are used to predict the next arrival station and arrival time of the vehicle.The vehicle's actual and predicted arrival station and arrival time are compared to perform vehicle identity authentication.Experimental results show that the proposed scheme can use the centralized global control capabilities of Software Defined Network（SDN）and the distributed computing capabilities of MEC to manage and assign vehicle authentication tasks.The scheme ensures a high authentication accuracy of vehicles and satisfies the low latency requirements of IoV.

【Key words】Internet of Vehicles（IoV）；authentication；behavior prediction；Software Defined Network（SDN）；Mobile Edge Computing（MEC）

DOI：10. 19678/j. issn. 1000-3428. 0056614

0 概述

近年来，随着物联网、自动驾驶技术的快速发展，车联网（Internet of Vehicles，IoV）相关领域的研究进展也受到了人们的广泛关注。车联网作为物联网的一个重要分支，主要为城市交通环境中的驾驶员、乘客和交通管理人员提供网络接入[1]，实现车辆与车辆、车辆与行人、车辆与道路基础设施及车辆与云平台之间的互联互通。然而，由于这些交互过程均以网络为载体，因此车联网通信环境自身的脆弱

基金项目：国家自然科学基金（61972080）；上海市青年科技启明星计划（19QA1400300）；同济大学嵌入式系统与服务计算教育部重点实验室开放课题（ESSCKF2019-01）。

作者简介：杨雪婷（1997—），女，硕士研究生，主研方向为车联网安全；李 重（通信作者），副教授。

收稿日期：2019-11-15 **修回日期：**2020-01-27 **E-mail：**yangxt@mail.dhu.edu.cn

性和信息安全问题也阻碍了其进一步的发展，例如黑客通过盗窃车载设备、渗透车辆内部网络以及利用外部网络对车辆实施攻击，然后通过这些被攻击的异常车辆对车联网环境中的其他用户进行干扰，从而严重损害用户利益甚至威胁用户人身安全。可见，保证车联网安全至关重要，并且由于网络安全依托网络中各成员的行为规范，因此保证车联网安全的前提是确定所有入网车辆的身份合法性。

在移动无线网络中，传统身份认证技术主要是基于密码学的身份认证技术，例如：文献[2]提出使用基于公钥基础结构的传统公钥加密方案，其向每个参与者发布一个绑定其身份的证书和公钥，但当参与者数量增加时，证书管理会变得非常困难；文献[3]提出基于身份的公钥加密技术，其验证过程简单，不需要证书。在此基础上，文献[4]提出基于双线性配对的身份公钥加密技术，文献[5]基于椭圆曲线密码学来验证移动用户的身份，该方法在保证安全性的情况下密钥长度较短。在车联网中，多数身份认证技术也主要是基于密码学的认证技术，例如：文献[6]使用匿名公钥和私钥对进行车辆身份认证，并为每辆车提供匿名证书；文献[7]提出基于智能卡协议的匿名轻量级身份验证方法，使用低成本的加密操作来验证车辆和数据消息的合法性；文献[8]利用一次性公钥构造时间戳的签名，并将其作为认证信息实现匿名认证；文献[9]提出一种高效的基于身份的聚合签名方案，通过批量验证思想提高验证效率。

上述身份认证方案虽然可满足车辆身份认证的基本要求，但其作为一种静态防御机制仍存在两方面的问题。一方面，传统身份认证技术无法解决车辆身份盗用问题。身份盗用是指正常车辆的车载设备被非法操控者盗窃使车辆行为出现异常，或者黑客通过无线网络攻击直接控制车辆的行为使车辆异常，这两种情况都会使车联网安全环境受到威胁。这时利用传统身份认证方案仅能识别其身份，不能及时检测出车辆是否存在异常行为。另一方面，随着5G时代的到来，网络架构将会发生巨大变化，未来5G网络架构将会是小单元部署和覆盖的异构结构，这种较小的单元部署和高速的车辆移动特性使用户和接入点之间的认证交互变得更频繁[10]。在此情况下，需要更快、更高效的交互认证满足用户的低时延需求，然而基于密码学的身份认证方案多数需要复杂的计算，这将严重影响用户体验。本文在移动边缘计算（Mobile Edge Computing，MEC）框架下，提出一种基于车辆行为预测的身份认证方案，利用软件定义网络（Software Defined Network，SDN）的集中式全局控制能力和移动边缘计算的分布式计算能力，并根据车辆的历史行为数据对车辆未来行驶行为进行预测，从而实现车辆身份认证。

1 网络架构

由于本文研究车联网领域的身份认证技术，因此首先应确定车联网的网络架构。在5G技术的引领下，目前较主流的网络架构是MEC和SDN结合的网络架构[11-13]。SDN被认为是一种利用提供流可编程性和网络弹性来优化5G网络管理的新技术[14]，通过将控制平面与数据平面分离，把转发规则的控制权交由SDN控制器，实现高效和智能的数据转发以及对网络管理的全局控制。MEC作为云计算的扩展，以分布式部署应用程序和服务的方式在接近移动用户的一侧提供计算能力，其通过将计算任务从中心云转移到移动边缘设备来缓解网络压力。因此，将两者结合形成的MEC+SDN新型网络架构具有可靠、低时延、可扩展等特性。

目前，车联网中主流的身份认证机制是以可信第三方为基础，利用公钥加密等技术实现，如公钥基础设施通过网络介质建立实体之间的身份信任关系，验证终端是否掌握正确的口令或密钥，实现对权限的掌控。换言之，车联网身份认证问题需要一个可信的全局控制中心。软件定义网络将网络的控制功能与转发功能解耦，能为网络管理提供全局状态信息，因此本文利用SDN控制器实现车辆身份认证。但由于道路环境中等待认证的车辆规模庞大且身份认证机制存在的高通信负载、计算量和存储开销问题将影响认证交互的时效性，难以适应车辆节点高动态变化的特点，在车联网环境中车辆若要享受网络服务，则需要与基站或路侧单元等道路基础设施进行连接，而移动边缘计算的思想是在接近移动用户的一侧提供计算能力，因此本文将道路基础设施作为移动边缘计算站点来缓解SDN控制器在身份认证过程中的计算压力。综上所述，本文需要解决车联网身份认证问题，且判别车辆是否发生身份盗用需要大量的数据传输和计算工作，同时要满足用户的低时延需求，而MEC+SDN网络架构可用于解决车联网身份认证问题。

基于此，本文提出一种基于MEC框架的软件定义车联网体系结构，如图1所示。该体系结构将网络架构分为控制平面和数据平面，在体系结构的顶层（即控制平面）中设置一个SDN控制器，可对底层设备（如交换机）进行全局管理，灵活控制网络流量。中间层是由交换机组成的核心网络，通过有线链路提供数据转发和状态采集功能。本文将基站或路侧单元等道路基础设施作为MEC服务站，通过网络功能虚拟化来调度、存储和计算资源。同时，每个MEC服务站可看作是负责与车辆通信的SDN子控制器。架构的底层是车辆之间的无线通信网络，使用IEEE 802.11p通信协议。车辆身份盗用通常发生在车间通信或车辆与MEC服务站之间的通信过程中。针对车联网身份认证问题，

本文利用SDN控制器的集中控制能力对车辆的身份认证任务进行指导,同时将计算任务下放到中间层MEC服务站实现对底层车辆的安全认证并达到降低认证时延的目的。

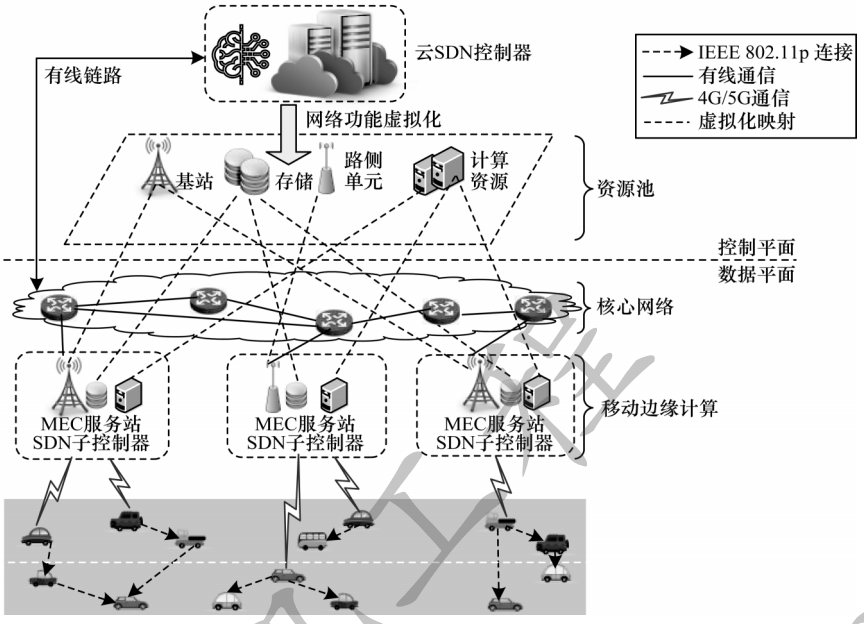

图1　基于MEC框架的软件定义车联网体系结构

Fig.1　Architecture of software defined IoV based on MEC framework

2　相关定义

2.1　车辆身份盗用

分析车辆身份盗用发生的可能性首先需要了解车载电子设备的实际攻击面。根据文献[15]所述,车载电子设备的安全威胁主要分为3类,分别为车载设备的盗窃、内部网络的渗透以及外部网络的利用。车载设备的盗窃是指音/视频播放器、导航仪、数字多媒体广播接收器、远程信息处理设备等被盗。内部网络的渗透是指攻击者将恶意设备连接到汽车电子控制单元以进入车内网络来控制车辆,该攻击可以控制制动器或关闭发动机等造成的人身伤害和车辆严重损坏。另外,由于车辆内部网络使用CAN协议,而该协议存在一些安全漏洞,因此利用CAN总线的攻击方式也层出不穷,例如读取、泛洪和重放攻击等,而这些攻击方式能在保证车辆通过身份认证的情况下控制车辆行为。利用外部网络的攻击是由车辆无线连接设备滥用外部网络引起,例如在车间通信或车辆与道路基础设施的通信过程中,攻击者通过信道攻击实现窃听等损害用户利益的行为。

通过分析车载电子设备可能面临的安全威胁可得出,车辆身份盗用通常发生在车载设备的盗窃、内部网络的渗透或外部网络的利用这3种情况下,攻击者在车辆通过身份认证进入车联网环境后,通过网络漏洞入侵车辆使其行为异常,例如利用该异常车辆向其他车辆散播虚假道路信息或直接控制该异常车辆的行为导致事故的发生。

2.2　车辆行为

由于车辆在驾驶行为上存在时空分布规律性[16],每一部车辆都有其特定的行为模式,一般车辆行为包含时间、位置、速度和方向夹角,由于方向夹角与道路实际部署情况有关,因此本文仅考虑时间、位置和速度3种行为特征。时空分布规律是指一辆正常的车辆以某一时间、速度到达固定的地点后,通常会以另一个固定的时间到达另一个固定的地点,例如某上班族如果在工作日的07:00以40.2 km/h的速度从家里出发,则目的地通常是公司且到达时间稳定在08:00。因此,本文考虑利用时间、位置和速度这3个特征的历史数据挖掘潜在的行驶规律来预测车辆未来到达的位置和时间。使用一个五元组来表示预测车辆行为的数据:$H_i=(\mathrm{date},t,\mathrm{lon},\mathrm{lat},v)_i$,其中,$H$表示车辆行驶状态,$i$表示车辆标识,date表示当前日期,$t$表示当前时间戳,lon表示车辆所在位置的经度,lat表示车辆所在位置的纬度,v表示车辆行驶速度。

2.3　历史行为数据

本文设置基站上报机制,在基站完成对车辆的认证后,将认证情况和车辆行为数据上报至SDN控制器,以便SDN能掌握车辆的历史行驶轨迹并收集大量的历史行为数据。基站上报信息格式为:$(\mathrm{date},t,\mathrm{lon},\mathrm{lat},v,0/1)_i$,其中,前5项为车辆行驶状态,最后一项表示对车辆认证的判定结果,0表示车辆异常,1表示车辆正常。历史行为数据是指SDN控制器只要一收到基站对车辆的行为数据就上报,并将

其进行存储,形成该基站标识下该车辆的历史行为数据库。

2.4 网络初始化

本文假设所有基站均真实可信,且在网络初始化时采用传统基于密码学的认证方案对车辆进行认证,使SDN能掌握车辆的历史行驶轨迹和大量的历史行为数据。在本文认证方案中,SDN控制器掌握整个应用场景的地图和基站的部署情况,并负责对车辆认证任务进行分配。基站负责执行SDN控制器分配的计算任务并对车辆进行行为认证。此外,车辆的行驶轨迹使用途中接入的基站相连接表示,并将车辆在路径起始位置处连接的基站称为该轨迹的起始点,例如车辆i从家里出发去公司,依次接入了基站BS1、BS2、BS3、BS4、BS5这5个站点,在接入BS1之前没有基站向SDN上报车辆i的认证结果,那么这条路径表示为BS1→BS2→BS3→BS4→BS5,起始点为BS1。基站首先在路径的起始点对车辆进行基于密码学的认证来验证其身份,然后在后续认证中只对其行为进行认证。

3 基于车辆行为预测的认证算法设计

3.1 算法简介及具体步骤

传统基于密码学的认证方案虽然能满足车辆身份认证的基本要求,但无法解决车辆身份盗用问题,即车辆身份标识和密码被敌手所盗取,而第三方认证机构无法辨识车辆真实身份的问题,同时车辆的高速移动特性使身份认证交互在小单元部署的网络架构中变得更加频繁,而基于密码学的认证方案计算复杂度高,不能满足用户的低时延需求。为此,本文在基于MEC框架的软件定义车联网体系结构下,创新地提出一种基于车辆行为预测的认证方案,该方案不仅能解决车辆身份盗用问题,而且能满足用户的低时延需求。在基于车辆行为预测的认证方案中,将SDN集中式控制能力和MEC分布式计算能力相结合。首先,SDN控制器通过分析车辆历史行为数据,提前通知MEC站点做认证准备工作,当车辆到达MEC站点通信范围内时可快速对车辆进行认证,这种提前将认证计算任务分配给MEC服务站的机制不仅能够降低SDN控制器的计算压力,而且能实现对车辆的快速认证,从而解决车联网认证时延问题。其次,考虑车辆可能发生身份盗用问题,本文利用车辆历史行为数据对车辆未来行为进行预测,通过真实和预测行为的对比来判断车辆是否出现行为异常,从而解决车联网中的车辆身份盗用问题。

在本文基于车辆行为预测的认证框架中,SDN控制器主要负责收集并匹配车辆的历史行为数据,根据历史行为数据确定认证基站,并将对应的历史数据发送至认证基站,通过认证基站对车辆行为进行预测,这样计算任务就下发至中间层MEC服务站,MEC服务站主要负责对车辆行为进行预测并在车辆到达时对比真实行为和预测行为,实现对底层车辆的安全认证,最终将车辆行为数据上发至SDN控制器。本文通过SDN和MEC的协同作用实现对车辆的认证,并达到降低认证时延的目的,具体认证步骤如下:

步骤1 假设车辆i通过基站j的认证后,基站j将车辆的行驶状态和认证结果上报给SDN控制器,SDN根据其掌握的车辆i的历史行驶轨迹进行搜索,基于前缀树确定可能的下一个站点(即一个或多个执行认证准备的认证基站),并将相应的历史行为数据发送给认证基站。

步骤2 被SDN控制器选择的认证基站收到车辆i的历史行为数据后,基于决策树和多元非线性回归方法对车辆i即将到达的下一个站点和到达时间进行预测。

步骤3 当车辆行驶到预测到达站点的通信范围内并请求认证时,该认证基站只需将预测时间与车辆真实的到达时间进行对比,并设置一个误差阈值α:若不超过阈值α,则表明车辆i正常,认证通过;若车辆i不在该时间阈值范围内到达,则认为其行为出现异常,认证不通过。

步骤4 认证基站将车辆i的认证情况上报至SDN控制器。

3.2 基于前缀树的认证基站确定

SDN控制器需要在上一个站点上报车辆认证结果后根据历史数据搜索可能的下一个站点,即一个或多个做认证准备的认证基站,并将车辆历史行为数据发送给这些认证基站进行车辆行为预测和认证。理论上,车辆到达上一个站点后,根据实际道路部署情况,车辆可能到达的站点是附近所有可通过道路直达的站点,因此应该通知上一个站点周围所有的基站。但由于车辆在驾驶行为上存在时空分布规律性,车辆特定的行为模式使得车辆行驶轨迹存在关联性,即已知车辆此次行驶的部分轨迹后,车辆接下来经过的站点也有规律可循。如果能大致找出车辆经过的下一个站点并通知其做认证准备,并排除车辆不会到达的站点,则可大幅降低车联网认证的计算负载。

前缀树是一种树形结构,主要用于统计、排序和保存字符串,尤其是保存关联数组,其优势在于查询效率高。前缀树将数据库中出现的项严格按字典顺序排列,使出现的项组成一个偏序集。考虑车辆行驶轨迹的关联性,即始点和终点确定时,整条路径基本趋于稳定,本文设计使SDN控制器根据基站上报的车辆行驶数据构建车辆轨迹前缀树,将车辆历史行驶轨迹上的每一个站点作为一个项,组成偏序集,当已知车辆此次行驶过的站点时,沿着前缀树搜寻下一个站点,可大幅缩小认证基站的选择范围,明显

提高计算效率。在车辆行驶过程中,SDN控制器可根据车辆行驶过的站点在前缀树中搜索可能的下一个站点,该方式相比广泛通知附近所有站点做认证准备要更加节约通信成本和计算资源。

本文基于如图2所示的示例场景,进行基于前缀树的认证基站确定。假设SDN控制器保留的数据库中车辆ID1的历史行驶轨迹如下:

BS1→BS4→BS7→BS10→BS12→BS14→BS16

BS1→BS4→BS7→BS10→BS13→BS15

BS1→BS4→BS7→BS5→BS8

BS1→BS4→BS6→BS9→BS11

BS1→BS4→BS2→BS5→BS3

BS11→BS9→BS7→BS4→BS1

BS11→BS14→BS17→BS15

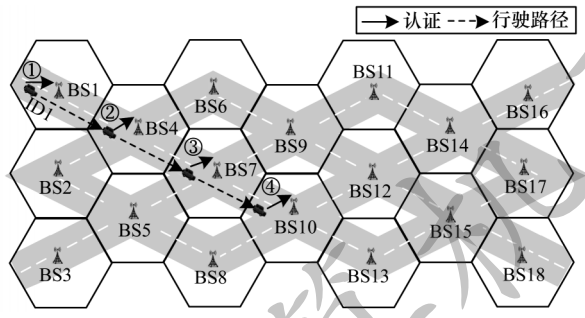

图2　示例场景

Fig.2　Example scenario

SDN控制器根据以上历史轨迹构建如图3所示的前缀树。该前缀树是以ID1为根的包含2个频繁项BS1和BS11的完全前缀树,表示2个频繁项所构成的所有可能的轨迹模式,每个子树表示以该子树的根为前缀的所有模式。当车辆ID1以BS1为起点出发时,SDN控制器只需搜索以BS1为根的子树,若车辆ID1以图2所示的路径行驶到BS10,其中①、②、③、④表示车辆ID1接入基站的顺序,则SDN控制器根据已行驶的轨迹BS1→BS4→BS7→BS10在前缀树中展开搜索,搜索方式如下:

步骤1　从根节点开始搜索。

步骤2　取得已行驶轨迹上的第1个站点BS1,根据该站点选择对应的子树并转到该子树继续进行搜索。

步骤3　在相应的子树上,取得要查找轨迹上的第2个站点BS4,进一步选择对应的子树进行搜索。

步骤4　依次迭代到站点BS10。

步骤5　取得已行驶的轨迹BS1→BS4→BS7→BS10中的所有站点后,读取附在BS10节点上的信息,得到结果为BS12和BS13,因此SDN控制器选择BS12和BS13作为下一步进行认证准备的认证基站,即完成认证基站的确定。

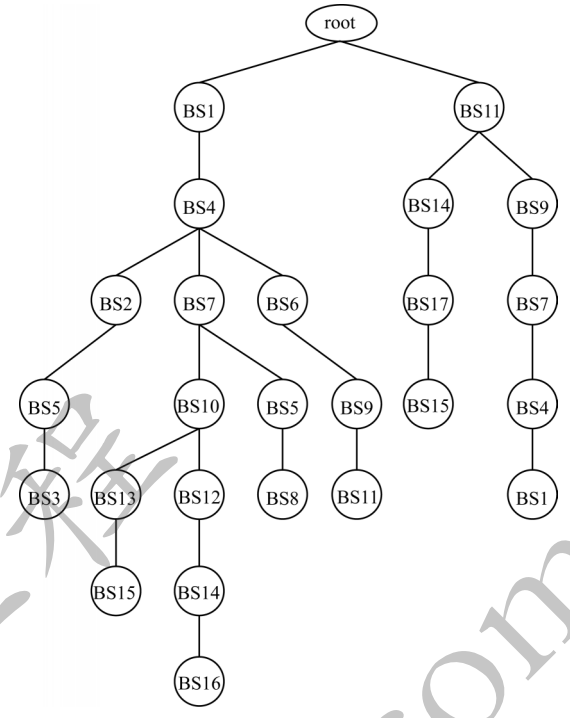

图3　车辆历史轨迹前缀树

Fig.3　Prefix tree of vehicle historical tracks

3.3　基于历史行为数据的车辆行驶状态预测

由上文分析可知,SDN控制器确定了做认证准备的认证基站后,会将车辆的历史数据发送给这些选定的认证基站,使其进行车辆行驶状态的预测。

3.3.1　数据格式预测

由网络初始化的前提可知,认证基站每结束对一辆车的认证都会将车辆行驶状态和认证结果上报给SDN控制器,SDN根据前缀树选择下一个或多个认证基站并在分配计算任务后等待认证基站的上报结果。当有认证基站上报时,SDN会根据上一个认证基站和当前上报的认证基站的数据合成预测数据对。下文结合图2的示例场景和图3的车辆历史轨迹前缀树举例展示SDN控制器的数据格式。假设车辆ID1从BS1出发,以$(2019.06.01,09:30:22,121.237\,962,34.225\,910,40.1)_{ID1}$的状态行驶到BS4并在此处完成认证,SDN控制器收到BS4的上报结果后,根据历史轨迹前缀树通知BS2、BS6和BS7做认证准备,假设车辆以$(2019.06.01,09:52:21,121.237\,962,34.312\,915,35.1)_{ID1}$的状态行驶到BS7,SDN控制器将在路径BS4→BS7关于车辆ID1的历史数据中添加一条数据对:$(2019.06.01,09:30:22,121.237\,962,34.225\,910,40.1)_{ID1}$→$(2019.06.01,09:52:21,121.237\,962,34.312\,915,35.1)_{ID1}$。

在路径BS4→BS2和BS4→BS6关于车辆ID1的历史数据中各添加一条数据对:$(2019.06.01,09:30:22,121.237\,962,34.225\,910,40.1)_{ID1}$→0,其中,0表示车辆未到达。假设网络初始化完成,SDN控制器已

掌握路径BS4→BS7关于车辆ID1的历史数据，当车辆ID1以状态（2019.06.04，08：29：15，121.237 962，34.225 869，29.1）$_{ID1}$到达BS4时，SDN发送给BS7关于车辆ID1的历史数据格式如图4所示。

图4 BS4→BS7数据对示例

Fig.4 Example of BS4→BS7 data pair

3.3.2 车辆行驶状态预测

由图4所示的BS4→BS7数据对可以看出，已知上一个站点的历史行为数据和对应的下一个站点的历史行为数据，当输入车辆当前在上一个站点的行驶数据时，要求根据历史数据和当前输入预测车辆即将到达下一个站点的行驶状态。本文考虑位置和时间两个因素先预测车辆将到达哪一个站点，再预测到达该站点的时间。

常用的预测算法及模型有神经网络、决策树、线性回归、支持向量机等，由于神经网络预测模型需要大量的训练数据且较长的训练时间，并且在现实交通场景中，车辆行驶路径与时间属性密切相关，例如在工作日的上班高峰期驾驶人通常以相同的路径出发去公司，因此对于车辆下一个站点的预测，本文选择能映射对象属性和对象值关系的决策树预测算法。对于到达时间的预测：一方面，由于时间属性是连续值且车辆行为认证需要实现细粒度的时间预测，而决策树算法只能将连续属性离散化进行预测；另一方面，车辆到达时间与交通环境的拥堵状态有关，很明显不呈线性关系，因此本文考虑利用车辆到达上一个站点的时间和速度两个因素选用多元非线性回归预测模型对到达下一站点的时间进行预测。

典型的决策树算法包括ID3、C4.5、CART等。ID3算法优先划分信息增益较大的特征，其更倾向于划分有较多属性值的特征。C4.5算法用信息增益率作为选择分支的准则，能够避免ID3算法的缺陷，但ID3和C4.5算法根据属性值划分数据，之后特征节点不再起作用，并且该快速划分方式会影响算法准确率。CART算法采用二元划分法，使用基尼指数选择最优的数据分割特征，是应用较为广泛的决策树算法，因此本文利用CART算法实现对车辆下一个站点的预测。

基于图4所示的BS4→BS7数据对示例，当前做认证准备的认证基站只需要判定车辆以某种状态到达上一个站点时是否会到达本站点，因此决策树的叶节点只有车辆到达或者未到达两种结果，即在做认证准备的认证基站侧的历史数据中，有车辆行驶状态的归类为车辆到达，没有车辆行驶状态（标记为0）归类为未到达。由于决策树叶节点的结果与到达上一个站点的状态有关，因此上一个站点的时间特征将作为决策树的分支属性。

CART算法使用基尼指数选择划分属性。基尼指数反映了从数据集中随机抽取两个样本，其类别标记不一致的概率。由于数据集的基尼值越小，纯度越高，因此CART算法选择使划分后基尼指数最小的属性作为最优划分属性。CART算法的具体过程为：首先确定建立决策树的特征，其次计算各个特征的基尼指数，然后选择使得基尼指数最小的特征进行划分。

对于特征选取，考虑到实际交通流量随时间呈现出的差异，本文首先选择“工作日标签0/1”作为一个特征，其中，1表示工作日，0表示非工作日，其次由于一天内不同的时间点呈现出不同的交通流量，这也将影响用户的路径选择，因此将时间戳作为另一个特征。

对于以上的时间戳属性，决策树不能直接处理连续值，因此需要先对时间属性进行离散化，本文根据实际生活的通勤时间将一天划分为00：00—06：00、06：00—10：00、10：00—16：00、16：00—19：00、19：00—24：00这5个时间区间。选取当前输入时间所在的大区间，再将其按30 min一个区间划分为多个小区间，并且根据实际预测精度的需要还可对该大区间进行更细致的划分，然后仅保留该大区间中的历史数据进行建树，这样不仅可减少计算量，而且能提高预测准确率。

假设认证基站保留的数据集为D，第k类样本所占比例为$p_k(k=1,2)$，数据集的纯度用基尼指数进行度量，具体计算公式为：

$$\mathrm{Gini}(D)=\sum_{k=1}^{2}\sum_{k'\neq k}p_k p_{k'} \tag{1}$$

假定离散属性a有W个可能的取值，若使用a对样本集D进行划分将会产生W个分支节点，其中第w个分支节点包含D中所有在属性a上取值为a^w的样本，记为D^w，因此属性a的基尼指数计算公式为：

$$\mathrm{Gini\_index}(D,a)=\sum_{w=1}^{W}\frac{|D^w|}{|D|}\mathrm{Gini}(D^w) \tag{2}$$

最终选择使得划分后基尼指数最小的属性作为最优划分属性[17]。在建立决策树后,将上一个站点的当前状态输入到决策树中,认证基站将预测出车辆是否会达到其覆盖范围:若到达,则进行下一步的到达时间预测;若未到达,则终止对该车辆的认证准备。考虑到车辆行驶的波动性,可能有一个或多个站点预测车辆即将到达,此时多个站点同时做下一步的到达时间预测。

由于本文需要实现细粒度的到达时间预测且交通道路复杂,车辆不是一直保持匀速行驶状态,其时间上不满足线性关系,因此本文采用多元非线性回归模型对车辆到达时间进行预测,将车辆到达上一个站点的时间和速度分别作为自变量 t^{pre} 和自变量 v^{pre},车辆到达下一个站点的时间作为因变量 t^{next}。首先选用不同的曲线估计模型对两个自变量和因变量的关系进行曲线拟合,然后通过拟合优度检验方法和回归方程的显著性检验方法在众多回归模型中找到简单且拟合效果好的曲线,最后综合两条曲线构建多元非线性回归模型,实现对车辆到达时间的预测。

常见的曲线估计模型如表1所示,其中 x 表示本文涉及的两个自变量中的任意一个,本文首先选用7种曲线估计模型对车辆到达上一个站点的时间 t^{pre} 和速度 v^{pre} 两个自变量和到达下一个站点的时间 t^{next} 之间的关系分别进行曲线拟合,利用最小二乘法求出各曲线的参数,然后检验各曲线的拟合优度。

表1　曲线估计模型

Table 1　The curve estimation model

模型名称	模型表达式
一元线性函数	$t^{\mathrm{next}}=a_0+a_1x$
二次函数	$t^{\mathrm{next}}=a_0+a_1x+a_2x^2$
三次函数	$t^{\mathrm{next}}=a_0+a_1x+a_2x^2+a_3x^3$
指数函数	$t^{\mathrm{next}}=a_0\mathrm{e}^{a_1x}$
逆函数	$t^{\mathrm{next}}=a_0+a_1/x$
幂函数	$t^{\mathrm{next}}=a_0x^{b_1}$
逻辑函数	$t^{\mathrm{next}}=(1/u+a_0a_1x)^{-1}$

拟合优度检验以判定系数为衡量标准,判定系数数值上等于相关系数的平方,记为 R^2,其计算公式为:

$$R^2=\frac{\sum_q(t_q^{\mathrm{next}}-\hat{t}_q^{\mathrm{next}})^2}{\sum_q(t_q^{\mathrm{next}}-\bar{t}^{\mathrm{next}})^2} \tag{3}$$

其中,t_q^{next} 为预测所用的第 q 条历史数据到达下一个站点的时间的实际值,$\bar{t}^{\mathrm{next}}$ 为实际到达时间的平均值,$\hat{t}_q^{\mathrm{next}}$ 为曲线估计值,当判定系数 R^2 越接近1时,模型拟合优度越高。本文联合假设检验和检验 F 对建立的回归方程进行显著性检验,F 统计量计算公式为:

$$F=\frac{\sum_q(\hat{t}_q^{\mathrm{next}}-\bar{t}^{\mathrm{next}})^2/m}{\sum_q(t_q^{\mathrm{next}}-\hat{t}_q^{\mathrm{next}})^2/(n-m-1)} \tag{4}$$

其中,n 为做预测时所用的历史行为数据的总数,m 为自变量的个数,由于本文是在构建多元非线性回归模型之前分别对到达上一个站点的时间和速度两个自变量进行参数估计,因此 m 取1。F 为统计量服从第一自由度为 m、第二自由度为 $(n-m-1)$ 的 F 分布,即 $F\sim(m,n-m-1)$。F 值越小,说明自变量对因变量造成的影响越小于随机因素对因变量造成的影响。F 值越大,代表回归方程的拟合优度越高[18]。

对于自变量 t^{pre} 和自变量 v^{pre},在各自的曲线拟合结果中选择判定系数较大且 F 统计量较小的曲线,将两条最优的拟合曲线相加,重新设置判定系数来构建最终的多元非线性回归模型,然后通过最小二乘法求解判定系数得到最终的多元非线性回归方程。SDN控制器将车辆在上一个站点的行为数据发送给当前认证基站后,认证基站将行为数据中的时间和速度两个数值代入到上述所求的多元非线性回归方程中,求得预测车辆到达当前站点的时间。

3.4　基于行为的车辆身份认证判定

当车辆到达预测的下一个站点的通信范围内时,该基站将车辆行为数据中的时间与预测的到达时间做差值:若不超过预设的时间误差 α,则认为车辆正常;若超过或者车辆到达一个未对该车辆做认证准备的站点,则认为车辆异常。最后基站将认证结果上报至SDN控制器。由于车辆行驶存在一定的波动性,因此若在基于决策树的到达站点预测中有多个站点预测结果为到达,则这些站点都将对该车辆进行认证,其中只有车辆唯一到达的站点对该车辆的认证结果为1(即车辆正常),其他站点将向SDN控制器上报车辆异常。此时,SDN控制器综合多个站点的上报情况,只要收到1个正常上报就认为车辆正常。

4　实验与结果分析

为验证本文认证方案的可靠性和有效性,本节将从认证准确率和认证时延两个方面对其进行性能评估。本文将上海市松江区的OSM地图导入SUMO交通仿真工具中生成模拟地图。所选实验区域如图5的方框所示,相应的仿真地图如图6所示。利用SUMO模拟50天的交通场景,生成车辆历史行为数据。模拟场景内车辆总数为3 400,所有车辆的轨迹随机分布,车辆行驶速度设置为0 km/h~60 km/h,采用 $N(28\,800,720^2)$ 和 $N(61\,200,720^2)$ 正态分布函数使车辆进入地图的时间形成07:00—09:00和16:00—18:00两个早晚高峰时间段。将时间从00:00开始换算成秒表示,正态分布函数自变量是车辆进入地

图的时间，因变量为车辆在当前时刻进入地图的概率。正态分布的第1个参数为期望，28 800表示车辆进入地图的平均时间为第28 800秒（即08:00），61 200表示车辆进入地图的平均时间为第61 200秒（即17:00）。第2个参数为方差，720^2表示在第28 800秒和第61 200秒时车辆进入地图的概率最大，为$1/(720\sqrt{2\pi})$。本文选取车辆ID truck 107在图5所示的行驶路径上的历史行为数据验证本文认证方案的有效性。

图5 实验区域

Fig.5 Experimental area

图6 仿真地图

Fig.6 Simulation map

4.1 基于多元非线性回归的到达时间预测分析

基于SUMO生成的车辆行驶数据，选取车辆“ID truck 107”在行驶路径中上一个站点和下一个站点的历史行为数据进行到达时间预测分析，其中上一个站点的经纬度为121.152 699°和31.000 765°，下一个站点的经纬度为121.182 256°和31.003 878°。将上一个站点的时间和速度两个因素作为自变量，车辆到达下一个站点的时间作为因变量，分别对两个因素进行参数估计，从而实现模型选择。

4.1.1 曲线拟合

到达上一个站点时间与到达下一个站点时间的曲线拟合结果如表2所示。从判定系数R^2和F值两项结果可以看出，在拟合到达上一个站点时间与到达下一个站点时间的曲线时，二次函数的R^2为0.872，相对于其他函数的判定系数较大，且F值为160.258，相对于其他函数较小，显著性F变化量（Sig.）为0，小于0.005，这说明用二次函数进行到达上一个站点的时间与到达下一个站点的时间的曲线拟合效果较好，形式如下：

$$t^{\text{next}} = a_0 + a_1 \cdot t^{\text{pre}} + a_2 \cdot (t^{\text{pre}})^2 \tag{5}$$

其中，t^{pre}表示车辆到达上一个站点的时间，t^{next}表示车辆到达下一个站点的时间，a_0为常数项，a_1、a_2为二次方程的系数。

表2 到达上一个站点时间与到达下一个站点时间的曲线拟合结果

Table 2 Curve fitting results of the arrival time to the previous station and the arrival time to the next station

模型名称	R^2	F	Sig.	参数估计值			
				a_0	a_1	a_2	a_3
一元线性函数	0.867	314.161	0	−23 920.402	2.121		
二次函数	0.872	160.258	0	−180 900.355	15.064	0	
三次函数	0.872	160.258	0	−180 900.355	15.064	0	0
指数函数	0.863	303.590	0	4 194.311	7.75E−5		
逆函数	0.870	321.580	0	79 057.418	−1.25E+9		
幂函数	0.865	308.802	0	0.000	1.883		
逻辑函数	0.863	303.590	0	0.000	1.000		

对到达上一个站点的速度与到达下一个站点的时间的曲线拟合结果如表3所示。可以看出，二次函数的判定系数R^2为0.852，相对于其他函数的判定系数较大，且F值为136.452，相对于其他函数较小，Sig.为0，小于0.005，这说明用二次函数进行到达上一个站点的速度与到达下一个站点的时间的曲线拟合效果较好，形式如下：

$$t^{\text{next}} = b_0 + b_1 \cdot v^{\text{pre}} + b_2 \cdot (v^{\text{pre}})^2 \tag{6}$$

其中，v^{pre}表示车辆到达上一个站点的速度，b_0为常数项，b_1、b_2为二次方程的系数。

表3 到达上一个站点速度与到达下一个站点时间的曲线拟合结果

Table 3 Curve fitting results of the arrival speed to the previous station and the arrival time to the next station

模型名称	R^2	F	Sig.	参数估计值			
				b_0	b_1	b_2	b_3
一元线性函数	0.816	213.553	0	43 717.05	−319.42		
二次函数	0.853	136.452	0	8 880.765	1 074.86	−13.877	
三次函数	0.853	136.875	0	2.06E+4	373.55	0.000	−0.91
指数函数	0.814	210.674	0	49 716.54	−0.012		
逆函数	0.776	166.016	0	12 249.04	7.71E+5		
幂函数	0.794	184.598	0	2.63E+5	−0.576		
逻辑函数	0.814	210.674	0	2.011E−5	1.012		

4.1.2 到达下一个站点的时间预测模型

根据曲线拟合结果可以构建出以到达上一个站点的时间和速度为自变量、到达下一个站点的时间为因变量的多元非线性回归方程，形式如下：

$$t^{\text{next}} = a_1 \cdot t^{\text{pre}} + a_2 \cdot (t^{\text{pre}})^2 + b_1 \cdot v^{\text{pre}} + b_2 \cdot (v^{\text{pre}})^2 + c \tag{7}$$

其中，c为常数项。利用Python编程实现非线性回归，最终得到以到达上一个站点的时间和速度为自变量、到达下一个站点的时间为因变量的多元非线性回归模型，形式如下：

$$t^{\text{next}} = -2.507 \cdot t^{\text{pre}} + 7.743 \times 10^{-5} \cdot (t^{\text{pre}})^2 + 193.967 \cdot v^{\text{pre}} - 3.722 \cdot (v^{\text{pre}})^2 + 42\,513.53 \tag{8}$$

本文将到达下一个站点的真实时间与预测模型所得的预测时间（将时间从00:00开始换算成秒表示）进行对比，如图7所示，其中，横坐标轴表示仿真模拟天数，纵坐标轴表示到达时间，可看出该多元非线性回归模型拟合效果较好，适用于到达站点的时间预测。

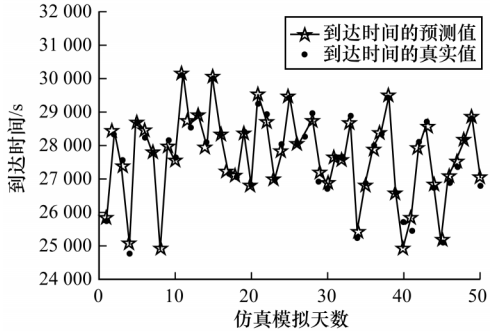

图7　到达时间的真实值与预测值对比

Fig.7　Comparison between the real value and the predicted value of arrival time

4.2　认证时延分析

本文选用文献[19]提出的VGKM和文献[20]提出的PPDAS两种车联网中基于密码学的双向认证方案进行实验对比，从认证时延的角度验证本文认证方案的低时延性能。本文在图5所示的行驶轨迹上选取多个站点，对在该路径上行驶的100辆车进行认证，测试并得出本文认证方案中将车辆真实行为与预测行为进行对比的平均认证时延为0.13 ms。如果在认证基站覆盖范围内有n辆车等待认证，则最大认证时延不超过$(0.13\times n)$ms。

3种认证方案的认证时延如图8所示，其中，横坐标轴表示同时在一个认证基站上等待认证的车辆数量，纵坐标轴表示平均认证时延。从图8可以看出，3种方案的认证时延随着车辆数量的增加而增加，当车辆数量达到100时，本文认证方案的认证时延比VGKM方案的认证时延少64.9%，比PPDAS方案的认证时延少67.5%。因此，本文认证方案在降低认证时延方面具有更好的性能优势。

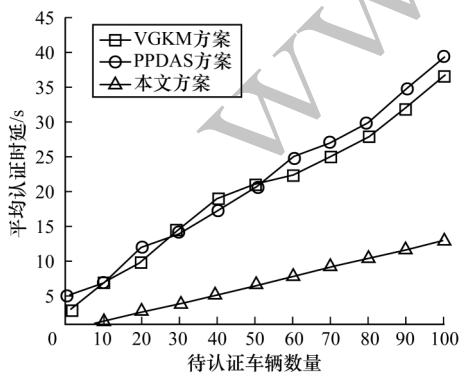

图8　认证时延对比

Fig.8　Comparison of authentication delay

4.3　认证准确率分析

为验证本文认证方案的安全性，在已知车辆即将到达的下一个站点及到达时间的情况下，实验分别引入不同比例的异常车辆，在到达时间误差设置在2 min和1 min两种情况下，通过选取的多个站点分别对100辆车的行为进行预测以验证本文认证方案的认证准确率，如图9所示，其中，横坐标轴表示异常车辆的比例，纵坐标轴表示认证车辆的准确率。可以看出，不论引入多大比例的异常车辆，在到达时间误差设置为2 min时，对车辆认证的准确率能保持在96%以上，当对认证精度要求比较严格而将误差设置在1 min时，对车辆认证的准确率略微下降但仍能保持在93%以上。实验结果表明，本文认证方案能在允许一定差错率的情况下保障车联网安全。

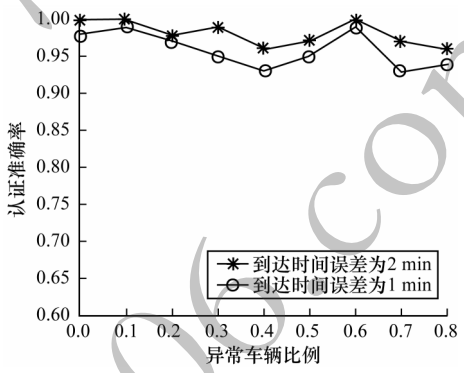

图9　认证准确率对比

Fig.9　Comparison of authentication accuracy

5　结束语

本文分析车联网中身份认证的安全性和低时延需求，结合MEC框架和软件定义车联网，提出一种基于车辆行为预测的身份认证方案。SDN控制器根据车辆历史行驶轨迹提前通知车辆可能到达的下一个站点进行认证准备，认证基站依据车辆历史行为数据预测准确的到达站点和到达时间，待车辆驶入预测的认证基站的覆盖范围内时，该认证基站只需对比车辆真实到达时间与预测到达时间即可实现车辆身份认证。在基于MEC框架的软件定义车联网体系结构混合架构中，每一个MEC站点分布式执行对车辆的认证工作，从而降低SDN控制器的负载并满足认证的低时延要求。实验结果表明，该方案在认证准确率和认证时延方面均具有较好的性能表现。但由于车联网中的车辆可能存在传感器故障或用户散布虚假道路信息等情况，此时仅使用车辆身份认证技术无法保证车辆信息交互安全，因此后续将设计一种快速准确的信任评估方案，使得车辆在接收消息前需先进行信任评估，进一步提升车联网的安全性。

参考文献

[1] CHENG Jiujun, CHENG Junlu, ZHOU Mengchu, et al. Routing in Internet of vehicles: a review [J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(5): 2339-2352.

[2] MENEZES A J, OORSCHOT P C, VANSTONE S A. Handbook of applied cryptography [M]. [S. l.]: CRC Press, 2018.

[3] SHAMIR A. Identity-based cryptosystems and signature schemes[M]. Berlin, Germany: Springer, 1985.

[4] HE D J, CHEN C, CHAN S, et al. Secure and efficient handover authentication based on bilinear pairing functions[J]. IEEE Transactions on Wireless Communications, 2012, 11(1): 48-53.

[5] CAO Jin, MA Maode, LI Hui. An uniform handover authentication between E-UTRAN and non-3GPP access networks[J]. IEEE Transactions on Wireless Communications, 2012, 11(10): 3644-3650.

[6] WANG Nengwen, HUANG Yuehmin, CHEN Weiming. A novel secure communication scheme in vehicular ad hoc networks [J]. Computer Communications, 2008, 31(12): 2827-2837.

[7] YING B D, NAYAK A. Anonymous and lightweight authentication for secure vehicular networks [J]. IEEE Transactions on Vehicular Technology, 2017, 66(12): 10626-10636.

[8] HUO Shiwei, YANG Wenjing, HOU Yintao, et al. Security analysis and improvement of anonymous authentication scheme in vehicle ad hoc network [J]. Computer Engineering, 2018, 44(5): 124-127. (in Chinese)
霍士伟, 杨文静, 侯银涛, 等. 车载自组网匿名认证方案的安全性分析与改进[J]. 计算机工程, 2018, 44(5): 124-127.

[9] LÜ Liudi, ZHENG Dong, ZHANG Yinghui, et al. Identity-based aggregation signature verification in vehicular ad hoc network [J]. Computer Engineering and Design, 2018, 39(7): 1866-1871. (in Chinese)
吕柳迪, 郑东, 张应辉, 等. 车联网中基于身份的聚合签名认证[J]. 计算机工程与设计, 2018, 39(7): 1866-1871.

[10] DUAN Xiaoyu, WANG Xianbin. Authentication handover and privacy protection in 5G HetNets using software-defined networking[J]. IEEE Communications Magazine, 2015, 53(4): 28-35.

[11] WANG Kai, YIN Hao, QUAN Wei, et al. Enabling collaborative edge computing for software defined vehicular networks [J]. IEEE Network, 2018, 32(5): 112-117.

[12] HUANG Xumin, YU Rong, KANG Jiawen, et al. Exploring mobile edge computing for 5G-enabled software defined vehicular networks[J]. IEEE Wireless Communications, 2017, 24(6): 55-63.

[13] LIU Jianqi, WAN Jiafu, ZENG Bi, et al. A scalable and quick response software defined vehicular network assisted by mobile edge computing [J]. IEEE Communications Magazine, 2017, 55(7): 94-100.

[14] AGIWAL M, ROY A, SAXENA N. Next generation 5G wireless networks: a comprehensive survey [J]. IEEE Communications Surveys and Tutorials, 2016, 18(3): 1617-1655.

[15] KOSCHER K, CZESKIS A, ROESNER F, et al. Experimental security analysis of a modern automobile[C]// Proceedings of 2010 IEEE Symposium on Security and Privacy. Washington D. C., USA: IEEE Press, 2010: 1-7.

[16] YAN Yang, SUN Lijun, ZHU Lanting. Short-term traffic flow prediction method based on spatiotemporal relativity[J]. Computer Engineering, 2020, 46(1): 31-37. (in Chinese)
闫杨, 孙丽珺, 朱兰婷. 基于时空相关性的短时交通流量预测方法[J]. 计算机工程, 2020, 46(1): 31-37.

[17] ZHOU Zhihua. Machine learning[M]. Beijing: Tsinghua University Press, 2016. (in Chinese)
周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.

[18] STEVEN M K. Fundamentals of statistical signal processing—estimation and detection theory[M]. Beijing: Electronic Industry Press, 2011. (in Chinese)
STEVEN M K. 统计信号处理基础——估计与检测理论[M]. 北京: 电子工业出版社, 2011.

[19] VIJAYAKUMAR O, AZEES M, KANNAN A, et al. Dual authentication and key management techniques for secure data transmission in vehicular ad hoc networks[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(4): 1015-1028.

[20] LIU Yanbing, WANG Yuhang, CHANG Guanghui. Efficient privacy-preserving dual authentication and key agreement scheme for secure V2V communications in an IoV paradigm [J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(10): 2740-2749.

编辑 陆燕菲