隐私保护的网约出行的研究综述

于海宁[1], 张宏莉[1], 余翔湛[1], 曲家兴[2]

[1] 哈尔滨工业大学网络空间安全学院 哈尔滨 中国 150001

[2] 黑龙江省网络空间研究中心 哈尔滨 中国 150001

摘要 为高效利用交通资源, 在线网约出行(ORH)服务整合车辆供给和乘客请求信息, 派遣符合条件的车辆提供非巡游的出行服务。人们在享受 ORH 服务带来的便利时, 也面临着严重的隐私泄露风险。为此, 许多研究利用密码学技术设计隐私保护的 ORH 服务。首先, 本文介绍了隐私保护的 ORH 服务主要面临的城市动态场景下高效计算密态行程开销、实时动态规划密态行程、安全共享不同 ORH 服务的运力资源等挑战。然后, 回顾了欧式距离、路网距离和行驶时间三类行程开销的安全计算方法, 其中, 欧式距离计算效率高, 但误差大, 现有路网距离和行驶时长的安全计算方法多数面向静态路网场景, 针对城市动态路网场景的安全计算方法有待进一步研究。分析了面向司机、乘客、ORH 平台的行程规划问题的求解方法, 现有研究往往仅针对司机、乘客或 ORH 平台的单一目标进行行程规划, 事实上行程规划不但要考虑 ORH 平台自身收益, 更要同时兼顾乘客和司机的用户体验。综述了隐私感知的行程预处理方法, 单车单客模式、单车多客模式的行程安全共享方法, 并总结了其不足与启示。多车单客、多车多客动态模式的行程安全共享有待进一步研究。最后, 从城市动态路网下高效的密态行程开销的安全计算与比较、多方隐私保护的大规模密态行程动态规划与安全保障、跨服务域的去中心化密态行程协作共享、ORH 服务的法律法规合规保证四方面展望了隐私保护的 ORH 服务的未来研究方向。本文旨在保护多方隐私的前提下, 提高 ORH 服务质量、促进多 ORH 服务合作, 使得网约出行更加智慧、更加安全。

关键词 位置隐私; 多方安全计算; 网约出行服务; 行程动态共享

中图法分类号 TP309.2 DOI 号 10.19363/J.cnki.cn10-1380/tn.2024.01.01

A Survey on Privacy-Preserving Ridesharing for Online Ride Hailing Services

YU Haining[1], ZHANG Hongli[1], YU Xiangzhan[1], QU Jiaxing[2]

[1] School of Cyberspace Science, Harbin Institute of Technology, Harbin 150001, China

[2] Heilongjiang Province Cyberspace Research Center, Harbin 150001, China

Abstract For efficient utilization of transportation resources, Online Ride Hailing (ORH) services enable riders to hail available vehicles by matching vehicle supplies and rider requests. Along with the advantage of ORH services raises serious privacy concerns. Thus, many studies focus on privacy-preserving ORH services by using some well-established cryptographic primitives. Firstly, we introduce main challenges for privacy-enhanced ORH services under in dynamic city scenarios, including encrypted travel cost computation, encrypted trips dynamic planning and secure ridesharing between different ORH services; then, we review secure travel cost computation about Euclidean distance, road distance and travel time. Euclidean distance secure computation is very efficient but not accurate. Most road distance and travel time secure computation methods are designed for static road networks, but it is desirable to have more methods for dynamic city road networks. We analyze ridesharing dynamic scheduling oriented from riders, drivers and ORH platforms. Existing works solve ridesharing dynamic scheduling with single optimal objective from riders, drivers or ORH platforms. Actually, comprehensive multiple objectives of riders, drivers and ORH platforms should be considered, such as income of ORH platforms, user experience of riders and drivers. We further summarize privacy-aware trip preprocessing and privacy-preserving ridesharing over encrypted trips, including single driver-single rider mode and single driver-multiple riders mode, and then further point out disadvantages and inspiration of existing studies. However, it is desirable to have secure ridesharing of multiple drivers - single rider dynamic mode and multiple drivers - multiple riders dynamic mode. Finally, we summary further works in privacy-preserving ORH services, including secure travel cost computation and comparison over encrypted trips, multiparty privacy-preserving dynamic ridesharing and safety guarantee over large scale encrypted trips, decentralized secure ridesharing cross ORH services, legal and regulatory compliance guarantee. The paper aims to improve the quality of an ORH service and enhance cooperation cross ORH services, while protecting the privacy of all parties. It can make ORH services more intelligent and more secure.

通讯作者: 于海宁, 博士, 副研究员, Email: yuhaining@hit.edu.cn。

本课题得到国家自然科学基金项目(No. 62172123, No. 61732022)和黑龙江省自然科学基金优秀青年项目(No. YQ2021F007)资助。

收稿日期: 2022-04-02; 修改日期: 2022-05-16; 定稿日期: 2023-09-26

Key words location privacy; secure multi-party computation; ride hailing service; dynamic ridesharing

1 引言

交通拥堵一直是困扰当代城市发展的主要难题之一, 其增加了人们的出行支出, 损害了人们的身心健康, 且带来了巨大的社会成本和经济损失, 例如, 时间价值损失、额外燃料消耗、环境污染加剧、交通事故增加等。交通信息分析公司 INRIX 的报告指出, 仅 2018 年美国由堵车造成的经济损失约为 870 亿美元。北京市交通发展研究中心发布的《2020 北京市交通发展年度报告》显示, 2019 年北京市日均拥堵时间超 6h。据百度公司测算, 北京市每年因交通拥堵造成数以千亿计的经济损失, 约占北京市 GDP 的 5%。治理交通拥堵已经成为全世界共同关注的问题, 从车辆端进行改革与优化是缓解交通拥堵最直接和最高效的途径之一。为此, 在移动互联网与普适计算的推动下, 大量在线网约出行服务(Online Ride Hailing, ORH)纷纷涌现, 例如, Uber, Lyft 和滴滴出行。ORH 服务是以移动互联网技术为依托构建的服务平台, 其整合车辆供给和乘客需求信息, 派遣符合条件的车辆提供非巡游的出行服务。如表 1 所示, ORH 支持多种模式的出行服务, 其中, 出行共享也被称为“共乘”、“拼车”、“行程共享”, 指多位起点/终点不同的乘客, 经过协商共同乘坐同一辆车并分担费用的共享出行方式。出行共享让有限的网约车供给和城市道路资源得到最大化利用, 为构建智慧、高效、绿色的出行生态提供巨大助力。出行共享因其社会和经济价值已得到广泛的关注, 用户规模也呈爆炸式增长。2019 年 11 月 29 日滴滴出行公司发布数据显示, 自 2015 年滴滴拼车上线以来累计有 29 亿人次使用, 最近一年累计行驶 45 亿 km, 平均每天为乘客节省等待时间 276 万 min。

人们在享受 ORH 服务带来的便利出行时, 面临着严峻的隐私泄露问题。为使用 ORH 服务, 用户需向平台提交其行程信息, 然而, 行程信息包含了大量的用户隐私, 例如, 用户的上下车位置和时段。获取这些隐私信息后, 攻击者能够实时地跟踪目标用户, 或者结合背景知识推断出用户的生活/工作地址、兴趣爱好和经济状况等隐私。在 ORH 服务中, 潜在的攻击者包括:半可信/恶意的乘客、司机、ORH 平台和外部敌手, 例如, ORH 平台为了获得额外广告推送收益, 收集用户行程信息用来推断用户的住址和兴趣爱好等; 司机为了跟踪特定的乘客, 伪造自身状态向 ORH 平台骗取乘客行程; ORH 平台为了赢得商业竞争, 作为外部敌手窃听或攻击其他 ORH 平台。

表 1 ORH 服务的出行服务模式

Table 1 Travel mode in ORH services

行程共享模式	实例	描述
单车单客动态模式[1]	滴滴快车/专车/出租车	车辆只允许为一名乘客服务, 车辆无预设行程
单车多客静态模式	滴滴顺风车	车辆允许与多名乘客共享行程, 车辆有预设行程且出发后不可调整
单车多客动态模式	滴滴拼车/出租车拼车	车辆允许与多名乘客共享行程, 车辆无预设行程且出发后可调整
多车单客静态模式	公交车	一名乘客分段地与多台车辆共享行程, 车辆有预设行程且出发后不可调整
多车单客动态模式	—	一名乘客分段地与多台车辆共享行程, 车辆无预设行程且出发后可调整
多车多客静态模式[2]	—	多台车辆与多名乘客共享行程, 车辆有预设行程且出发后不可调整
多车多客动态模式	—	多台车辆与多名乘客共享行程, 车辆无预设行程且出发后可调整

(注: 1.单车单客静态模式可视为一种特殊的单车多客静态出行共享模式, 故未单独列出。2.多车多客模式可视为多车单客模式的进一步扩展)

已有研究[1-4]揭示了目前主流的 ORH 服务都存在着不同程度隐私泄露风险, 这些隐私威胁不但会危害用户个人权益, 而且很可能因违反《数据安全法》和《个人信息保护法》等法律规定导致处罚, 损害平台声誉。针对这些隐私威胁, 可以运用隐私保护技术处理原始数据, 改变其形式, 进而避免原始数据中敏感信息的泄露。如图 1 所示, 为保护敏感信息, 原始数据可以被变换为去标识化数据、脱敏数据、匿名化数据或加密数据等形式, 一般来说, 虽然它们对应的处理开销递增, 但同时也达到了更高的隐私保护强度。相应地, ORH 服务的隐私保护解决方案可分为两类: 一类解决方案是使用 k-匿名, l-多样性、t-近似性或差分隐私等非加密的方法来保护用户隐私, 但这类方法往往仅能提供有限的隐私保护强度, 且引入的噪声数据会影响 ORH 服务质量, 导致行程安排存在误差或错误; 另一类解决方案是采用加密技术, 将乘客请求和车辆状态加密后上传至 ORH 平台, 数据以密文形式进行传输和存储, ORH 平台/外部敌手无法获取用户真实的行程信息, 即使 ORH 平台被攻击后造成数据泄露, 用户也不担心自身隐私

泄露。此外, 基于加密的方法提供了更高的隐私保护强度, 且避免了噪声数据添加对服务精度的影响。考虑到 ORH 服务对行程规划的高精度要求, 本文重点关注基于加密技术的隐私保护解决方案。

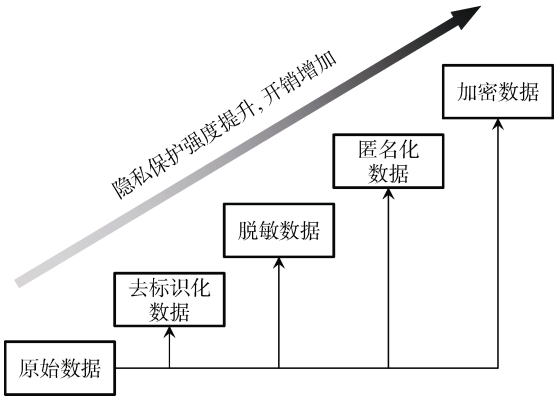

图 1 隐私保护下的数据形式

Figure 1 Data format with privacy preservation

本文通过全面总结分析了 ORH 服务行程动态共享与隐私保护的相关理论和技术, 力图探索如何在满足在 ORH 服务多方隐私保护需求的前提下, 最大化地利用 ORH 服务的运力资源和城市道路资源。本文主要的研究工作如下。

1) 总结分析了行程开销的安全计算方法, 具体包括欧式距离的安全计算、路网距离和行驶时长的安全计算。

2) 总结分析了行程共享的规划问题求解方法, 具体从面向司机、乘客、ORH 平台三方面分析各类规划目标和约束的求解方法。

3) 总结分析了隐私保护的行程共享方法, 具体包括隐私感知的行程预处理、单车单客模式和单车多客模式的行程安全共享。

本文结构组织如下: 第 2 节定义系统模型和服务模式; 第 3 节介绍构建隐私保护的 ORH 服务面临的挑战; 第 4 节总结分析行程开销的安全计算方法; 第 5 节总结分析行程共享的规划问题求解方法; 第 6 节总结分析隐私保护的行程共享方法; 第 7 节描述研究挑战与未来展望; 最后, 总结全文。

2 系统模型与服务模式

在 ORH 服务中, 用户主要包括乘客与司机, 他们通过 ORH 平台建立联系, 构成双边市场的供给与需求。如图 2 所示, ORH 服务的工作流程如下:

1) 请求提交: 乘客向 ORH 平台发送包含其行程信息的网约请求, 包括行程起点/终点、上车/下车时间窗口、最大容忍的绕路开销、乘客人数等。

2) 状态更新: 车辆以一定频率向 ORH 平台更新其状态, 包括当前位置、当前行程表、当前空闲座位等, 其中行程表记录了车辆未来的目的地(上客/落客地点)的时序安排, 以及到达各目的地容许的时间窗口。

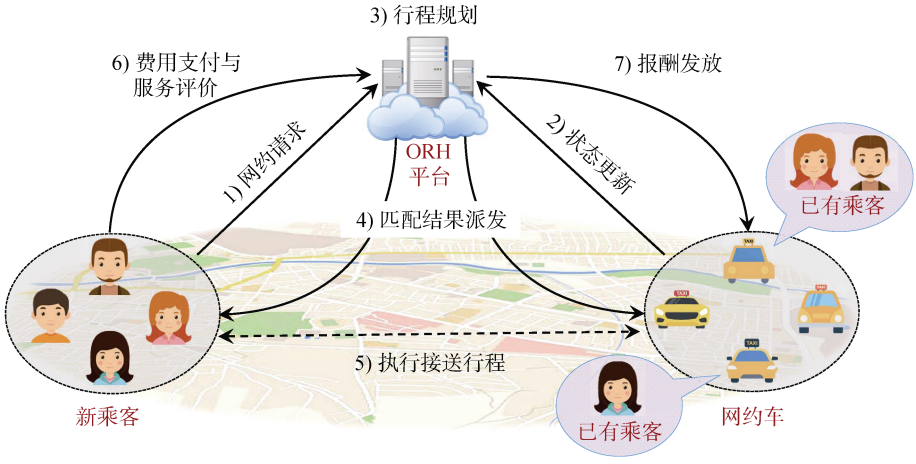

图 2 ORH 系统模型

Figure 2 System model of an ORH service

3) 行程规划: 基于乘客请求和车辆状态, ORH 平台将请求分配给相应车辆并调整车辆行程表, 使车辆在满足行程共享的约束条件下, 达成一定的优化目标, 并生成行程费用。

4) 匹配结果派发: ORH 平台将匹配结果发送给相关用户进行确认, 乘客收到司机和车辆信息、预计上车/下车时间、支付费用等, 司机收到乘客信息、新的行程表、酬劳等。

5) 接送行程执行: 经双方确认后, 司机按照新的行程表, 在容许的时间窗口内将乘客由起点位置

送达终点位置。

6) 费用支付与服务评价: 行程执行结束后, 乘客向 ORH 平台支付相应费用, 并对司机做出评价。

7) 报酬发放: ORH 平台扣除部分收益后, 向司机发放行程酬劳。

按照业务场景不同, 网约出行共享分为静态共享和动态共享两类:在行程静态共享中, 所有网约请求提前预知, 出发前一次性完成行程规划, 出发后车辆不能改变行程; 在行程动态共享中, 网约请求仅在发布时才能获知, 出发后仍可以重新动态规划行程, 随时调整车辆行程。本质上, 静态共享是动态共享在某一时刻收到所有网约请求的特殊情况。明显地, 行程动态共享能够合理利用车上的每一个座位, 最大化地利用有限的网约车供给和城市道路资源。本文关注的业务场景是城市动态场景下大规模网约出行动态共享, 即表 5 中的 1、3、5、7 项。

3 面临的挑战

采用加密技术实现隐私保护的行程动态共享仍面临着以下挑战:

1) 在城市动态场景下如何高效地计算密态行程开销?任意位置间的行程开销计算是 ORH 平台进行行程共享规划的基础依据。考虑到欧式距离密态计算的简易性, 多数隐私保护的 ORH 服务方案[2-3,5-6]用欧式距离度量行程开销, 但这忽视了车辆沿路网行驶这一关键特征。欧式距离无法准确地体现真实的行程开销, 例如, 使用欧式距离代替路网距离进行 KNN 查询存在约 25%的错误率[4]。计算密态路网距离本质上是基于加密图的最短路径计算问题, 现有安全最短路径算法[4,7-10]的效率难以支持大规模行程动态规划需求。在城市动态路网下, 基于密态的乘客请求和车辆状态实时计算行程开销以支持大规模行程动态规划是一个亟待解决的挑战问题。

2) 在城市动态场景下如何实时地动态规划密态行程?行程动态规划是 ORH 平台完成行程共享的核心环节, 除了单车单客的出行模式(这类问题求解相当于二分图匹配问题, 在多项式时间内可求最优解), 其他行程共享模式的规划问题已证明是 NP 难问题[11], 这类问题求解相当于扩展旅行商问题, 即多项式复杂程度的非确定性问题。因此, 求解行程共享的规划问题即使在明文下也是具有相当难度的, 例如, 滴滴拼车指出其每一次成功的行程共享平均需要额外的 18.6 万次计算, 每日处理数据超过 4875TB。目前, 多数隐私保护的 ORH 服务方案关注单车单客的行程动态匹配[2-6]或单车多客的行程静态规划问题的同态求解[12-16]。仅有 Yu 等[17]提出的 PSRide 首次实现了单车多客的行程动态规划问题的同态求解。针对单车单客、单车多客、多车单客、多车多客 4 类出行动态共享模式, 高效地同态求解行程共享的动态规划问题是一个极具难度的挑战问题。

3) 在城市动态场景下如何安全地共享不同 ORH 服务的运力资源?目前, 网约车市场处于群雄逐鹿的态势, 各公司纷纷推出各自的 ORH 服务, 例如, Uber, 滴滴出行、嘀嗒出行、首汽约车、曹操出行等。这些 ORH 服务拥有各自的用户群体和运力资源, 且互不连通共享。在城市出行高峰时段, 乘客经常面临着约车难的痛点。如果能够整合不同 ORH 服务的运力资源实现跨服务域的行程共享, 这将进一步提升运力资源的利率和用户体验。然而, ORH 服务提供商往往担心商业隐私泄露和支付纠纷, 使得弱信任的 ORH 服务之间难以实现跨服务域的行程共享。构建去中心化多 ORH 服务间的行程安全共享体系, 多 ORH 服务跨域协作完成行程安全动态共享是一个前瞻性的挑战问题。

4 行程开销的安全计算

城市路网下起点和终点之间的行程开销是指车辆由起点到终点所行驶的路网距离或时长, 每一次行程共享需要进行大量的行程开销计算, 以判断行程规划方案的可行性和优劣程度。因此, 密态行程开销的计算与比较开销决定了后续行程动态规划的效率。目前, 行程开销的安全计算主要包括: 欧式距离的安全计算、路网距离或行驶时长的安全计算。

4.1 欧式距离的安全计算

欧式距离的安全计算多利用同态加密算法[18-19]、多方安全计算协议[20-21]实现, 其计算方法已相对成熟, 且效率较高。Pham 等[3]采用类同态加密(SHE)算法[18]实现了多个欧式距离的并行同态计算; Liu 等[22]采用部分同态加密(PHE)算法[19]和姚氏协议[20]实现了欧式距离的安全计算; Bringer 等[23]采用扩展的茫然传输(OT)协议[24]实现了欧式距离的安全计算。上述算法最大的优势是效率高, 支持大规模并发运算, 但以欧式距离衡量行程开销忽视了车辆沿路网行驶这一关键特征, 使其无法准确地体现真实的行程开销。事实上, 欧式距离不宜作为行程开销的测量标准, Luo 等[4]和 Yu 等[17,25]通过真实的城市路网验证发现使用欧式距离代替路网距离进行 KNN 查询存在约 25%的错误率。

4.2 路网距离和行驶时长的安全计算

路网距离和行驶时长的安全计算本质上是图论

中最短路径的安全计算, 路网距离和行驶时长在速度已知的前提下可相互转换。最短路径查询作为图结构加密[9]中的基本操作, 可以利用可搜索加密(SSE)技术[26-27]、茫然数据结构[10,28]、多方安全计算(SMC)技术[29]实现最短路径安全计算。但这类方法的计算和通信开销极高, 无法适用于城市网路场景, 例如, Carter 等[30]提出基于姚氏协议[20]的最短路径算法在 100 个节点的路网中需几分钟计算一次最短路径; Liu 等[31]提出的基于 ORAM(Oblivious RAM)的最短路径算法在 1024 个节点的路网中计算一次最短路径需要 GB 级的通信量。

近年来, 针对大规模图结构, 已提出了一些效率提升的最短路径算法。Shen 等[8]使用类同态加密算法和保序加密(ORE)算法[32]在加密的图结构上实现了近似最短距离的安全查询, 但 ORE 仅能提供有限的隐私保护; Meng 等[7]使用类同态加密算法和距离预言机[33]实现了三种类型的近似最短距离安全查询算法, 但算法的准确率和效率有待提高; Liu 等[34]设计了一种新的图结构对称加密方案以支持高效的最短距离查询, 该方案构造 2HCL(2-hop cover labeling)[35], 以快速判断两顶点之间是否可达, 并求两顶点之间的近似距离; Luo 等[4]利用路网嵌入技术将路网嵌入到高维空间, 而后利用部分同态加密算法和姚氏协议在高维空间中计算近似的网路距离, 该方法需要在互不串通的双服务器场景下使用, 且服务器端的计算和通信开销较高; 针对文献[4]的不足, Yu 等[36]在路网嵌入的高维空间中利用一种改进的部分同态加密算法和密文双重致盲方法实现了单一服务器的近似最短路网距离安全计算, 且极大地降低了服务器端的计算和通信开销。上述方法实现了近似最短路网距离的安全计算, 其精度往往正比于计算和通信开销, 即当对结果精度要求较高时, 会带来更高的开销。为此, 一些研究聚焦于精确路网距离的安全计算。Wang 等[9]采用部分同态加密算法和姚氏协议实现了隐私保护的 Dijkstra’ s 最短路径算法, 并利用斐波那契堆来提高计算效率, 但该方案需要为用户部署可信代理协助计算; 在文献[34]的基础上, Zhang 等[37]提出了一种基于部分同态加密[38]的图加密方案, 以支持精准最短距离的安全计算。Yu 等[25]提出了基于超立方体路网嵌入的最短路网距离精确计算方法, 该方法的性能明显优于现有方法。然而, 上述方案也需要在互不串通的双服务器场景下使用。

上述最短路网距离安全计算方法主要存在两个局限性, 使其难以直接应用于城市动态路网下的密态行程动态规划: 1) 上述方法仅支持静态路网下的最短路网距离安全计算, 然而在动态路网下最短路网距离经常动态变化, 而这些方法难以支持路网的实时更新, 导致路网距离计算失效; 2) 上述方法在计算若干次最短路网距离时效率尚可, 但其开销难以满足行程动态规划中大规模最短路网距离实时计算的需求。

在隐私保护的实时导航领域, 一些安全导航方案[38-39]支持动态路网距离的安全计算。然而, 导航对路网距离计算的效率需求要远低于行程动态共享, 例如, Wu 等[39]利用私密信息检索(PIR)技术[29]和姚氏协议实现了一个隐私保护的实时导航协议, 在使用下一跳矩阵压缩技术提升效率后, 每更新一跳仍需要 1.5s 计算时长和 100KB 通信开销。过高的开销导致这些方案难以适用于城市场景下的行程动态共享。针对行程动态共享对行程开销计算的实时性需求, Yu 等[17]提出了一种基于区域锚点的最短行驶时长估计算法, 并利用部分同态加密算法和姚氏协议实现了密态行驶时长的安全计算, 但当路网的路况变化较快时, 该方法需要更多的通信开销保持算法准确性。

4.3 不足与启示

基于加密数据的行程开销计算能够有效地避免起止点位置隐私泄露。利用类同态加密技术及其密文打包技术, 欧式距离的安全计算效率极高, 一般情况下, 通过若干次同态乘法和加法能够同时计算出数千个欧式距离。基于加密数据的路网距离或行驶时长的计算更为复杂, 开销更大, 其通常需要利用路网嵌入技术将二维空间中无法密态计算的问题在高维空间中密态求解, 但其效率能够支持大规模的 KNN 密态实时查询操作。欧式距离难以有效地体现真实的行程开销。目前, 多数路网距离/行驶时长的安全计算方法仅面向静态路网场景。而针对动态路网场景, 能够支持行程动态共享的路网距离/行驶时长的安全计算方面的相关研究成果较少。针对动态路网距离/行驶时长的安全计算, 可将复杂的计算问题分解为可同态计算的简单问题, 并通过构建安全的索引结构提升效率。

5 行程共享的规划问题求解

行程共享的规划问题求解是: 给定路网上的车辆集合和乘客请求集合, 将不同的乘客请求分配给不同的车辆并安排车辆行程表, 使车辆在满足共享的约束条件下, 达成一定的优化目标。

5.1 面向司机的行程规划

司机通常期望以最短的距离服务乘客请求, 或

在最短的时间完成乘客请求, 因此, 面向司机的行程规划常用的优化目标包括最小化车辆行驶总距离和最小化车辆最大完工时间。针对静态场景下的行程共享问题, 一些研究[44,46-47]提出了最小化车辆行驶总距离的近似算法。目前, 研究更关注动态场景, 多数研究[41,45,48]采用插队操作, 以最小化车辆行驶总距离为优化目标计算行程动态规划问题的近似解。Ma 等[41]提出一种基于枚举的贪婪插队算法, 针对新的乘客请求, 该算法不重新规划车辆行程表, 而是将新请求贪婪地插入到行驶距离增加最少的车辆行程表, 该算法提高了求解的实时性, 但损失了求解的最优性。Huang 等[48]设计了一种基于字典树的数据结构(活动树)记录车辆的可行路径和开销, 乘客请求可以递归地插入到活动树中, 而车辆每条可行的路径对应着活动树中由根节点到叶节点的一条路径, 车辆选择总距离最小的路径行驶。Santi 等[45]提出了一种基于批次处理的插队算法, 该算法将乘客请求聚合成组, 而后按批次地尽可能多地将每个组内的乘客请求插入到当前车辆行程表。Tong 等[49]提出了一种基于动态规划的插队算法,极大地降低了计算开销, 并进一步提出了基于线性时间插队操作的通用框架,大幅减少了冗余的最短路径查询。此外, 还有一些研究[50-51]关注最小化所有司机完成最后一个订单的时间, 例如, Ascheuer 等[50]提出的重新规划框架和暂时遗忘框架, 前者当新请求到来时重新规划车辆行程表, 而后者在积攒一定的请求后批次规划行程表; Feuerstein 等[51]也提出了一种类似于重新规划框架的算法。有效地将插队操作与时空索引结合, 是未来进一步提升行程规划效率的可行方法。

5.2 面向乘客的行程规划

乘客通常期望尽快到达终点, 此外, 还期望能够与彼此间友好的乘客/司机共享行程。因此,面向乘客的行程规划常用的优化目标包括最小化等待时间和最大化社会效用。Xu 等[52]提出了最小化乘客最大等待时间的插队算法, 通过动态规划将查询最优插队位置的操作转化为查询特定区间最小值的操作, 并采用动态区间的查询索引达到线性时间复杂度。针对最小化乘客平均等待时间的优化目标, Waisanen 等[53]提出了一种基于先到先服务的策略以及基于旅行销售商问题的行程规划算法; Alonso 等[54]提出了一种基于订单打包的数据结构(RTV 图), RTV 图中若干位乘客间相互组合可枚举构成所有可行路径, 每条可行路径与其候选车辆间连边, 而后将行程共享问题转化为整数规划问题进行求近似解。Kameswaran 等[55]调研发现社交关系在行程共享中的重要性, 例如, 乘客希望能够与性格相似的人同行, 此外, 有些乘客能够接受经过风景名胜所产生的绕路开销。考虑到乘客的出行体验, 近年的研究[56,57]开始关注乘客之间、司机与乘客之间的社交关系, 通过定义社会效用模型来表示是否适宜共享行程、是否具有相同偏好等社交关系。Cheng 等[56]在社会效用模型中同时考虑了乘客之间、乘客与司机之间的社交关系, 提出了以最大化社会效用为优化目标的行程静态规划问题, 并采用基于插队操作的启发式算法求解问题; Fu 等[57]提出了基于社会效用的 Top-k 行程静态规划问题, 即为每位乘客提供社会效用最大的 k 条候选路径, 由乘客根据自己的偏好进一步选择。遗憾的是上述两个研究仅适用于静态场景。有效度量且高效计算社会效用, 并将其引入到行程动态规划成为了新的挑战。

5.3 面向 ORH 平台的行程规划

ORH 平台通常关注乘客请求成功完成数量以及平台总盈利。因此, 面向 ORH 平台的行程规划常用的优化目标包括最大化平台的乘客请求完成数和最大化平台的总收入。由于受到时空约束和车辆数量的限制, ORH 平台无法保证完成全部乘客请求, 为此, 有研究[58-61]力图在满足约束条件下为尽可能多的乘客提供服务, 即最大化平台的乘客请求完成数。Eleiner 等[58]计算每辆车接送若干位乘客的路网距离矩阵, 然后利用匈牙利算法规划行程, 确定车辆接送乘客的先后顺序; Santos 等[59]采用了基于插队操作的局部搜索算法在最大化乘客请求完成数量的同时最小化乘客花费; Cici 等[60]证明了最大化乘客请求完成数量的优化目标实际等价于最小化所需车辆数量; Vazifeh 等[61]提出了一个基于交通网络的算法求解最小化所需车辆数量这一等价问题。从商业利益的角度来看, 最大化平台的总收入也是一个重要的优化目标[62-64]。在实际系统中, 平台的总收入被定义为所有乘客支付的订单价格总和减去平台支付给所有司机的报酬。Asghari 等[62]提出了一种基于分支定界的框架, 以最大化平台的总收入为目标递归地尝试为每个新的乘客请求找到最优司机; Zheng 等[63]和 Biswas 等[64]都采用基于二分图匹配的算法, 尝试把可共享的若干乘客请求进行聚类, 而后通过枚举的方法把聚类好的乘客请求分派给一辆车, 但这类算法并不适用于大规模路网场景。之前介绍的方案[49]提出的统一的优化目标既可以等价地转化为最大化平台的乘客请求完成数, 也能够等价地转化为最大化平台的总收入。合理的行程定价, 以及公平、有效的用户激励机制是有待进一步研究的问题。

5.4 不足与启示

现有研究往往仅针对司机、乘客或 ORH 平台的单一角度进行行程规划。然而, 作为一个双边市场, ORH 平台不但要考虑自身收益, 更需要同时兼顾乘客和司机的用户体验; 此外, 司机和乘客所处的动态情景也没有充分考虑。应综合考虑司机、乘客和 ORH 平台三方利益制定自适应的行程动态规划的优化目标; 行程动态规划问题的求解方法能够在密文下高效实现, 即基于密态的乘客请求和车辆状态对行程动态规划问题求近似解。

6 隐私保护的行程共享

如表 1 所示, 行程共享模式包括: 单车单客模式、单车多客模式、多车单客模式和多车多客模式。现有行程安全共享方法主要关注不同模式下的用户身份隐私、用户位置隐私以及 ORH 平台商业隐私。这些方法可以分非加密的共享方法和加密的共享方法。非加密的共享方法常用技术包括: 空间隐匿[2]、伪造虚假位置[65]、位置偏移与模糊[66]和差分隐私[67]等。这类方法力图均衡隐私保护强度与服务质量, 具有较高的效率, 但它们仅提供了有限的隐私保护水平, 且泛化或噪声机制会影响服务精度。相反地, 加密的共享方法能够提供更高的隐私保护水平, 且不影响服务精度, 但它们计算开销较大。这类方法常用技术包括: 私密信息检索(PIR)[68]、多方安全计算(SMC)[20]、私密交集计算(PSI)[21]、类同态加密(SHE)[18]和部分同态加密(PHE)[19]等。针对不同共享模式, 现有的行程安全共享方法综述如下。

6.1 隐私感知的行程预处理

行程预处理的主要思想是: 通过某些过滤条件删除无法完成行程共享的乘客请求或车辆, 或通过聚类分组方法聚合乘客请求, 以达到快速的剪枝行程规划问题候选解规模的目的。Coslovich 等[40]提出了一种基于预先计算可行解的方法预处理乘客请求, 该方法采用两阶段的插入算法来提升效率, 其中第一阶利用乘客请求的时间间隔来预先计算可行解, 第二阶段再利用这些可行解对新请求进行匹配, 但该方法仅适用于中小规模的行程规划问题。Ma 等[41]提出了一种基于动态空间索引的方法过滤车辆集合, 该方法使用网格划分网路为区块, 为每区块建立车辆的动态时空索引, 然后, 根据乘客起点和终点所在区块索引双向筛选邻近区块内能满足时间约束车辆的作为候选集。Luo 等[4]同样采用的区块索引的方式过滤车辆, 其每次选择乘客起点所在区块及其邻近的 8 个区块内的车辆作为候选集, 并利用乘客起点到车辆的欧式距离进一步剪枝候选集; Yu 等[42]提出了动态条带索引结构, 采用条带划分的方法对移动物体进行索引获得了较高的可扩展性和查询效率, 并实现了数据的分布式处理。Gidofalvi 等[43]提出了一种基于请求聚类分组的方法预处理乘客请求, 该方法利用流数据分组技术把具有近似距离的起点/终点进行聚类分组, 并实现了并行化处理。类似地, Ta 等[44]和 Santi 等[45]也采用了聚类分组的方法预处理乘客请求。上述行程预处理方法没有考虑乘客/车辆的身份和位置隐私泄露问题, 很可能导致用户隐私泄露, 例如, 在某些车辆稀少区块内, 敌手易于实现对车辆的关联与追踪。行程预处理方法需要支持隐私的动态度量, 以实现剪枝力度与隐私保护强度之间的均衡。

区域的划分决定了隐私保护强度和预处理剪枝的力度, 例如, 当区域面积过小或某区域兴趣地点较少时, 利用背景知识很可能会推理出用户在区域内的准确位置。为此, Yu 等[36]提出应动态感知不同区域划分的隐私保护强度, 在满足隐私保护强的前提下最优化区域划分。具体地, 假设用户 u 在区域 z 中, 其他相关概念如下: $A=(a_1,\cdots,a_n)$ 表示 z 中的兴趣地点, 其概率分布满足 $p(a_i)\in[0,1]$ 和 $\sum_{i=1}^{n}p(a_i)$。$B=(b_1,\cdots,b_m)$ 表示隐私保护后敌手能够观察到的 z 中兴趣地点, 其概率分布满足 $p(b_j)=1/m$。$C=(c_1,\cdots,c_l)$ 表示关于 z 中兴趣地点的背景知识, 其概率分布满足 $p(c_k)\in[0,1]$ 和 $\sum_{c=1}^{l}p(c_k)$。基于上述假设, z 的条件信息熵代表了隐私保护的强度, 其计算如下:

$$H_z(A\mid BC)=-\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{l}p(a_ib_jc_k)\log_2 p(a_i\mid b_jc_k).$$

$H_z(A\mid BC)$ 越大表示隐私保护强度越高。至此, 可以依据 $H_z(A\mid BC)$ 确定区域划分的合理性: 指定隐私保护强度 δ 和区域划分 $\mathcal{Z}=\{z_1,\cdots,z_N\}$, 如果 $\forall z_i\in\mathcal{Z},H_{z_i}(A\mid BC)\geqslant\delta$, 则 $\mathcal{Z}$ 是一个合理的区域划分。

6.2 单车单客模式的行程安全共享

PrivateRide[2]首次提出了 ORH 服务的隐私保护问题, 其利用轻量级的匿名证书技术保护用户的身份隐私, 利用空间隐匿技术保护用户的位置隐私, 利用盲签名技术保护用户的支付隐私, 但该方法存在服务精度低, 且仅为司机提供了有限的隐私保护强度。ORide[3]改进了 PrivateRide 以增强用户隐私保

护水平, 其采用 SHE 技术和密文打包技术计算乘客与司机之间的密态欧式距离, 并据此完成行程匹配, 但 ORide 会对乘客泄露司机的位置隐私。TRACE[5]采用四叉树数据结构和轻量级空间隐匿技术保护用户位置隐私和 ORH 平台的地图划分隐私。CoRide[6]针对多 ORH 服务场景下的行程共享问题, 基于区块链的车雾计算技术提出了一种行程协作共享方法, 该方法采用私密逼近测试[69]和 Zerocash[70]保护用户身份隐私、位置隐私和支付隐私, 以及 ORH 平台的商业隐私。此外, 还有一些研究[71-72]基于区块链构建了隐私保护的 ORH 服务, 但区块链交易验证的吞吐率往往会成为系统效率瓶颈。考虑到密文计算效率, 上述隐私保护方案均采用欧式距离作为测量标准进行行程匹配, 但欧式距离无法体现真实的行程开销, 这势必导致行程匹配错误。Luo 等[4]发现使用欧式距离代替路网距离进行单车单客行程匹配存在约 25% 的错误率, 为此, 他们首次提出了基于路网距离的单车单客动态行程安全匹配方法(pRide), 该方法首先利用路网嵌入技术将路网转化为高维空间中的嵌入路网, 然后, 在嵌入路网下采用 PHE 计算密态路网距离, 采用姚氏协议比较密态路网距离, 最后, 选出具有最小行驶距离的车辆作为匹配结果。然而, pRide 在服务器端的计算和通信开销较大, 难以支持大规模的并发乘客请求。Yu 等[36]提出了一个轻量级的隐私保护方法, 即 lpRide, 该方法使用翻转加解密操作的 PHE 和嵌入路网技术实现了密态路网距离的高效计算, 使用盲密文分布机制实现了密态路网距离的高效比较, lpRide 大幅降低了服务器端的开销, 支持大规模的并发乘客请求。鉴于现有路网距离安全计算方法无法满足行程匹配对路网距离计算的实时性需求, 目前基于路网距离的单车单客密态行程匹配方法并不多。现有的 pRide 和 lpRide 也只能利用近似的路网距离进行行程匹配, 基于精确路网距离的行程安全匹配方法亟待深入研究。

6.3 单车多客模式的行程安全共享

PrivatePool[12]使用 PHE 技术和 PSI 技术设计了基于逼近测试行程安全规划方法和基于轨迹重合计算的行程安全规划方法, 前者通过私密计算乘客行程起点/终点与车辆轨迹之间的最小欧氏距离进行行程规划, 后者通过私密匹配乘客行程与车辆轨迹之间的最大重叠进行行程规划。在 PrivatePool 的基础上, Pagnin 等[73]提出了具有高效率的 O-PrivatePool 协议, 并进一步提出了考虑时间约束的 TOPPool 协议。PRIS[13]使用 PHE 技术面向加密的乘客和车辆的时空行程, 以最大化行驶时长节约量为约束优化目标进行行程安全规划, 其中, 行驶时长节约量等于乘客行程与车辆行程的总驶时长与该共享行程时长的差值。SRide[14]采用 SHE 和 SMC 技术保护乘客和司机的位置隐私, 该方法首先利用时空网格过滤候选司机, 然后, 使用改进的 Priv-2SP-SP 协议[74]计算乘客行程与司机行程的相似性, 并以最小化车辆行驶总距离为优化目标进行行程规划。Sherif 等[15]将乘客和司机加密后的行程转换为词向量, 通过比较乘客与司机之间的时间词向量相似性和空间词向量相似性, 以最大化乘客请求完成数为优化目标进行行程规划。FICA[16]采用基于区块链的车雾计算技术设计了一个保护乘客和车辆位置隐私的行程共享系统, 该系统使用分布式计算单元 RSU 构建记录交易的私有链, 由 RSU 负责上传处理乘客请求, 并利用基于时空标签的位置私密逼近测试进行行程规划。此外, 还有一些研究[67,75-76]基于差分隐私技术进行单车多客的行程共享, 其主要思想是通过对原有位置数据添加噪声保证司机与乘客的时空敏感信息不被泄露, 然后基于添加噪声后的数据进行行程规划, 但噪声的引入会影响行程规划的精准度。Yu 等[77]提出了一种基于群组中心点的车辆选择算法, 支持在密文下对群组行程共享问题进行同态求解, 并证明了密态欧式距离能够在群组行程共享中衡量行驶距离。上述单车多客模式的行程安全共享方法主要存在两个局限性: 1) 这些方法本质上是行程静态安全共享方法, 即司机和乘客的行程提前预知, 且开始行程共享后, 直至本次共享行程完成前不能更改行程, 显然这些方法无法适用于城市动态场景; 2)这些方法的行程共享仅能发生在车辆提前预设的有限位置附近, 乘客上下车位置灵活性较差。上述局限性与实际应用中乘客和司机的意愿相违背, 因为司机期望接送更多的乘客, 而乘客期望自定义上下车地点。

针对上述不足, Yu 等[17]首次提出了一个隐私保护行程安全动态共享方法—PSRide, 其以最短路网行驶时间为依据, 以最小化车辆绕道时长为优化目标, 在密文下动态规划行程, 即车辆在行程开始后, 若有空座位仍可以参加新一轮的行程规划, 当匹配成功时, 车辆动态调整行程表为新乘客服务。Yu 等[17]提出在密文下动态规划行程需要考虑计算和通信开销的问题, 规划方法不宜涉及过于复杂的运算, 因此, PSRide 利用动态插队算法以最小化附加行程开销为优化目标进行密态行程的动态规划。

假设乘客 u 的请求表示为 (l_s,l_d,t_s,t_d), 其中 l_s,l_d,t_s,t_d 分别表示上车地点、下车地点、上车截止

时间、下车截止时间。针对 u 的请求, 需要找到一辆车为其服务。假设任一辆车 v_x 的当前状态为 (l_x, S_x), 其中 l_x 表示车辆当前位置, S_x 是车辆的行程表, 表示为 $S_x = \langle (p_x[i], EAT_x[i], DLT_x[i]) \rangle_{i=0}^{N_x-1}$, $p_x[i]$ 是车辆未来停车上客或落客的位置, $EAT_x[i]$ 表示预计到达 $p_x[i]$ 的时间, $DLT_x[i]$ 表示到达 $p_x[i]$ 的截止时间。判断 v_x 能否为 u 服务, 需要测试 u 的请求能否插入到 v_x 的行程表 S_x 中, 即找到请求 (l_s, l_d, t_s, t_d) 在 S_x 的插入位置 (i, j), 使得新乘客 u 和 v_x 上已有乘客都不会迟到。如图 3 所示, l_s 插入在 $p_x[i-1]$ 和 $p_x[i]$ 之间, l_d 插入在 $p_x[j-1]$ 和 $p_x[j]$ 之间。让 $APT_{x,i}$ 和 $EPT_{x,i}$ 分别表示在 l_s 附加上客时长和预计上客时间, 随后让 $ADT_{x,i,j}$ 和 $EDT_{x,i,j}$ 分别表示在 l_d 附加落客时长和预计落客时间。

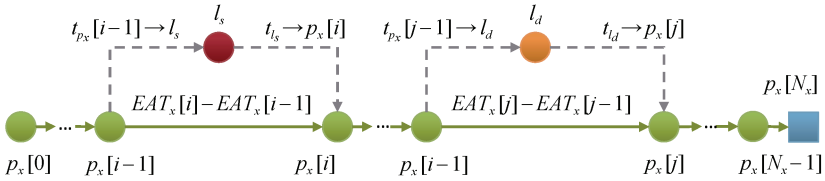

图 3 请求插入示例

Figure 3 An example of request insertion

通过时间维度上的加减运算可以直接计算出 $APT_{x,i}$, $ADT_{x,i,j}$, $EPT_{x,i}$ 和 $EDT_{x,i,j}$。根据这 4 个变量能够确定请求的插入车辆及其插入位置, 进而实现行程共享的动态规划。给定乘客 u 的请求 (l_s, l_d, t_s, t_d) 和车辆 v_x 的状态 (l_x, S_x), v_x 能够服务 u 当且仅当 $\exists (i,j), (0 \leqslant i \leqslant j \leqslant N_x)$ 满足如下时间约束:

$$APT_{x,i} \leqslant \min_{i \leqslant k < N_x} (DLT_x[k] - EAT_x[k]),$$

$$APT_{x,i} + ADT_{x,i,j} \leqslant \min_{j \leqslant k < N_x} (DLT_x[k] - EAT_x[k]),$$

$$EPT_{x,i} \leqslant t_s,\ EDT_{x,i,j} \leqslant t_d。$$

假设 $\mathcal{V}^*$ 表示能够服务 u 的车辆集合, 对于 $v_x \in \mathcal{V}^*$, FI_x 表示 v_x 可行的行程表插入位置集合。行程动态规划问题定义如下: 找到具有最小附加时间 ($ATT = APT + ADT$) 车辆 $v^* \in \mathcal{V}^*$ 服务 u, 即

$$(v^*, i^*, j^*) = \mathop{\arg\min}_{v_x \in \mathcal{V}^*, (i,j) \in FI_x} (APT_{x,i} + ADT_{x,i,j}).$$

该行程动态规划问题的求解过程仅涉及到加减法操作和比较操作, 因此, 基于前面介绍密态行驶时间的同态计算方法和密文安全比较方法, 采用部分同态加密算法和姚氏协议能够在密文下高效地求解上述规划问题。

目前, 城市动态场景下单车多客模式的行程安全动态共享方法较少, 尤其是以路网行驶开销为依据的行程安全动态共享方法亟待深入研究。

6.4 不足与启示

隐私感知的行程预处理能够在保证预期隐私保护强度的前提下最优化地剪枝用户请求或候选车辆。一般来说, 剪枝力度的增大会有效地降低开销, 但也可能在一定程度上会减弱隐私保护强度。因此, 可以根据敌手背景知识、隐私保护需求计算预处理对应的条件信息熵, 确定最优的剪枝策略, 进而降低开销。

行程安全共享能够有效地避免行程线路和时间表等隐私信息泄露。行程安全共享通常采用欧式距离、路网距离或行驶时间作为行程规划的基础依据。单车单客模式行程安全共享可以计算密态欧式距离进行行程规划, 其效率很高, 但会导致规划误差。在行程需处理的支持下, 单车单客模式行程安全共享应计算密态路网距离或行驶时间进行行程规划。目前密态路网距离或行驶时间的计算效率能够支持大规模实时的单车单客模式的行程安全共享。单车多客模式的行程安全共享无法采用欧式距离进行行程规划, 因其会导致行程规划错误, 即无法按时上客/落客。原则上, 单车多客模式的行程安全共享应采用路网距离或行驶时间作为动态规划的依据, 以避免规划错误。相对于单车单客模式, 单车多客模式的行程安全共享每一次行程规划需要更多的行程开销计算, 以及额外的时间表编排, 这对密文计算和通信效率提出了更高的要求。目前仅有少数研究[17]以行驶时间为依据才能够支持单车多客行程安全动态共享。

多数现有研究本质上是单车单客或单车多客模式的行程安全静态共享方法。一方面, 行程安全动态共享方面的研究成果较少; 另一方面, 多车单客、多车多客模式的行程安全共享方面尚待进一步研究。应重点关注单车多客、多车单客、多车多客模式的行程安全动态共享方法, 面向密态的乘客请求和车

辆状态, 构建多元化的优化目标, 在密文下高效地完成行程动态规划问题的求近似解。

7 研究挑战与未来展望

7.1 研究挑战

城市动态路网下大规模最小行程开销的安全计算问题: 针对现有在欧式空间或静态路网中密态行程开销计算方法难以适用于交通状况复杂的动态路网的困境, 在城市动态路网场景下, 如何高效地同态计算与比较密态行程开销, 以支持大规模密态最小行程开销的实时度量, 为后续密态行程动态规划提供决策依据, 这是要解决的基础性挑战问题。

ORH 服务域内的密态行程动态共享与安全保障问题: 针对城市动态路网下行程动态共享数据量大、动态性强、目标多样、模式灵活等特点, 在实时计算最小密态行程开销、社会效用和情景的基础上, 如何在单一 ORH 服务域内高效地完成多模式、多目标的密态行程动态规划, 并保护用户的身份隐私、位置隐私和支付隐私, 保障用户的人身安全[78-79], 这是要解决的核心挑战问题。

ORH 服务域间的密态行程协作共享与可控匿名支付问题: 针对多弱信任的 ORH 服务合作开展行程共享所带来的用户交叉认证需求和隐私跨服务域泄露风险, 如何去中心化地高效完成多 ORH 服务间的密态行程安全共享和流程合规监管, 以实现跨服务域的用户可控匿名认证、密态行程协作共享、费用可控匿名支付, 这是要解决的探索性挑战问题。

7.2 研究展望

7.2.1 城市动态路网下高效的密态行程开销的安全计算与比较

在城市动态路网场景下, 研究高效的密态行程开销的安全计算与比较方法, 以实现任意位置间的最短时空距离的实时安全度量, 为密态行程动态规划提供决策依据, 针对大规模动态路网下密态行程开销计算与比较存在的复杂度高、开销大等问题, 研究动态路网的高维空间嵌入机制, 以及动态嵌入路网中密态路网距离的同态计算方法, 研究动态路网下密态行驶时间的同态计算方法, 实现轻量、高效的密态行程开销的同态计算与比较, 使其在时空间维度上密态可度量; 针对密态行程开销同态计算的精准性和高效性兼顾的需求, 研究密态行程开销计算的准确率和效率之间量化机制, 为设备能耗、通信开销、计算效率、结果精度的优化均衡提供理论指导。

7.2.2 多方隐私保护的大规模密态行程动态规划与安全保障

在单一 ORH 服务场景下, 研究乘客、司机、ORH 平台三方隐私保护的大规模密态行程动态共享方法, 以实现面向多种共享模式的多目标优化的行程安全动态规划, 并保护用户身份、位置和支付隐私, 保障用户人身安全。

1) 支持隐私动态度量的行程预处理方法

针对隐私需求的主观敏感度个性化、场景差异化等特点, 在主客体隐私敏感度和应用场景智能感知的基础上, 研究融合主观认知和实时情景的隐私动态度量理论体系, 形成隐私风险、保护强度、泄露益损比、效能平衡的一致性量化方法和评价模型; 研究支持隐私动态度量的行程预处理方法, 包括基于动态空间安全索引的预处理方法, 基于行程茫然聚类分组的预处理方法, 在避免用户隐私泄露的前提下实现隐私可量、可控的行程快速筛选过滤, 以达到快速、安全地剪枝行程规划问题候选解规模的目的。

2) 融合社会效用与情景的密态行程多目标动态规划方法

针对城市动态场景下行程共享数据量大、动态性强、目标多样、模式灵活等特点, 研究多维度的行程动态规划的优化目标体系, 涉及面向司机的最小化行驶开销、面向乘客的最小化平均等待时间及最大化社会效用、面向 ORH 平台的最大化请求完成数及最大化总收入, 实现针对应用场景的多优化目标自适应融合; 研究社会效用与动态情景的相似度安全计算方法, 涉及用户社交关系、兴趣偏好、地点流行度、交通流量等因素的量化与同态计算, 以提高行程共享的用户体验和人文关怀; 研究支持高效同态计算的多模式密态行程动态规划方法, 面向单车单客、单车多客、多车单客、多车多客模式安全地动态规划行程, 具体包括可同态求解的动态规划插队算法、基于安全索引的行程加密搜索与动态剪枝机制、支持同态计算的行程规划启发式算法, 在满足社会效用、实时情景、时间窗口、绕路开销等约束条件下, 达成多优化目标的行程规划问题求近似解, 并保护用户身份隐私与位置隐私。

7.2.3 跨服务域的去中心化密态行程协作共享

在多个 ORH 服务共存的场景下, 研究跨 ORH 服务域的用户可控匿名认证机制[80-81]、密态行程协作共享协议、费用可控匿名支付机制[70,82-83], 实现弱信任的 ORH 服务间的行程跨域安全共享, 并保护各 ORH 服务的商业隐私、用户隐私和支付隐私。

1) 跨服务域的用户可控匿名认证机制

针对多ORH服务用户交叉认证存在的身份管理机制异构性、身份隐私跨域泄露风险、重复认证的额外开销等问题, 研究基于区块链的用户身份可控匿名认证机制, 在ORH服务域内构建基于匿名证书的用户身份匿名认证体系, 在ORH服务域间构建基于区块链和零知识证明的用户身份可控匿名认证体系, 在保证认证匿名性和可追踪性的前提下实现轻量级分布式的用户身份认证, 防止用户身份隐私泄露和匿名滥用。

2) 跨服务域的密态行程协作共享协议

针对ORH服务域间行程共享存在的约束优化目标异构、隐私跨域泄露等问题, 研究ORH服务域间的密态行程协作规划协议, 自适应地转换融合不同ORH平台的约束条件和优化目标, 在避免ORH平台的商业隐私泄露的前提下协作完成跨服务域的密态行程安全动态规划; 研究基于智能合约的跨域共享行程的自动化执行与监管机制, 探索面向行程跨域共享的智能合约形式化创建方法, 通过在区块链上构建多样性的智能合约, 自动化监督行程跨域共享执行的合规性和真实性。

3) 跨服务域的行程费用的可控匿名支付机制

针对跨ORH服务域支付的“前台匿名, 后台实名”的匿名性和可控性需求, 研究基于区块链的高效可控匿名支付机制, 用户支付时隐藏支付交易的来源、去向和金额, 监管部门执法时可追踪支付流程, 在保证高支付吞吐率的前提下实现支付的不可双花性、假名性、不可伪造性、可传递性、拆分聚合性和不可链接性(用户地址不可链接、多个支付不可链接、支付输入输出不可链接); 研究多方合作的穿透式支付监管策略, 将支付监管与用户身份管理、支付流程相结合, 实现对非法操作涉及的支付交易和参与方身份的追踪取证。

7.2.4 ORH服务的法律法规合规保证

随着《网络安全法》、《数据安全法》和《个人信息保护法》相继出台, 对服务提供商收集、处理数据应当尽到的合规要求也越来越明晰, 如果服务提供商在此过程中无法尽合规审查的义务, 可能会触发一系列相关法律法规的规定, 进而被追究刑事、民事和行政责任。数据收集需明确合法、正当、必要原则, 并明确收集的目的。即使收集方出于改善服务、数据分析等理由, 也应当谨慎处理收集信息的范围, 尤其是个人敏感信息。国家网信办、工业和信息化部、公安部、国家市场监督管理总局联合制定的《常见类型移动互联网应用程序必要个人信息范围规定》, 对ORH服务提供商收集用户个人信息范围进行了明, 具体包括: 1) 注册用户移动电话号码; 2) 乘车人出发地、到达地、位置信息、行踪轨迹; 3) 支付时间、支付金额、支付渠道等支付信息。虽然上述最少必要原则已尽可能地减少了ORH服务中隐私泄露风险, 但已有研究[1-4]揭示: 一方面, ORH服务可能存在潜在安全风险, 导致用户隐私泄露; 另一方面, ORH服务提供商结合背景知识和历史数据仍能够从最少必要收集的信息中推理出大量的敏感信息, 例如, 敏感地点、用户隐私等。为有效应对上述隐私泄露隐患, 应运用隐私保护技术对最少必要收集的信息进行进一步处理。具体地, 研究采用可控匿名认证技术[80-81]避免注册用户移动电话号码的泄露; 研究采用同态加密[18-19]、多方安全计算[20-21]、差分隐私[67]、匿名化[65-66]等隐私计算技术避免乘车人出发地、到达地、位置信息、行踪轨迹等行程信息的泄露; 研究采用可控匿名支付技术[70,82-83]避免支付时间、支付金额、支付渠道等支付信息的泄露。

8 结论

目前全球对ORH服务中用户隐私泄露的监管已愈发严格, 在保护用户隐私的前提下最大化地利用有限的运力资源已成为前沿的热点和难点问题。基于密码学的隐私保护技术提供了强隐私保护和高数据可用性, 已广泛地应用在ORH服务中。首先, 本文介绍了隐私保护的ORH服务面临的主要挑战; 然后, 分析了密态行程开销的安全计算、行程共享的规划问题求解方法和隐私保护的行程共享方法, 总结不足和启示; 最后, 展望了隐私保护的ORH服务的未来研究方向。本文旨在保护多方隐私的前提下, 提高ORH平台服务质量, 使得网约出行更加智慧、更加安全。

参考文献

[1] Zhao Q C, Zuo C S, Pellegrino G, et al. Geo-Locating Drivers: A Study of Sensitive Data Leakage in Ride-Hailing Services[C]. *Proceedings 2019 Network and Distributed System Security Symposium*, 2019: 1-15.

[2] Pham A, Dacosta I, Jacot B, et al. PrivateRide: A privacy-enhanced ride-hailing service[J]. *Proceedings on Privacy Enhancing Technologies*, 2017, 2017(2): 38-56.

[3] Pham A, Dacosta I, Endignoux G, et al. ORide: A Privacy-Preserving yet Accountable Ride-Hailing Service[C]. *The 26th USENIX Conference on Security Symposium*, 2017: 1235-1252.

[4] Luo Y C, Jia X H, Fu S J, et al. PRide: Privacy-Preserving Ride Matching over Road Networks for Online Ride-Hailing Service[J].

IEEE Transactions on Information Forensics and Security, 2019, 14(7): 1791-1802.

[5] Wang F W, Zhu H, Liu X M, et al. Efficient and Privacy-Preserving Dynamic Spatial Query Scheme for Ride-Hailing Services[J]. *IEEE Transactions on Vehicular Technology*, 2018, 67(11): 11084-11097.

[6] Li M, Zhu L, Lin X. CoRide: A privacy-preserving collaborative-ride hailing service using blockchain-assisted vehicular fog computing[C]. *International Conference on Security and Privacy in Communication Systems*, 2019: 408-422.

[7] Meng X, Kamara S, Nissim K, et al. GRECS: Graph encrytion for approximate shortest distance queries[C]. *The 22nd ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2015: 504-517.

[8] Shen M, Ma B L, Zhu L H, et al. Cloud-Based Approximate Constrained Shortest Distance Queries over Encrypted Graphs with Privacy Protection[J]. *IEEE Transactions on Information Forensics and Security*, 2018, 13(4): 940-953.

[9] Wang Q, Ren K, Du M, et al. SecGDB: Graph encryption for exact shortest distance queries with efficient updates[C]. *International Conference on Financial Cryptography and Data Security*, 2017: 79-97.

[10] Keller M, Scholl P. Efficient, oblivious data structures for MPC[C]. *International Conference on the Theory and Application of Cryptology and Information Security*, 2014: 506-525.

[11] Xu Y, Tong Y X, Li W. Recent Progress in Large-Scale Ridesharing Algorithms[J]. *Journal of Computer Research and Development*, 2020, 57(1): 32-52.
(徐毅, 童咏昕, 李未. 大规模拼车算法研究进展[J]. *计算机研究与发展*, 2020, 57(1): 32-52.)

[12] Hallgren P, Orlandi C, Sabelfeld A. PrivatePool: Privacy-Preserving Ridesharing[C]. *2017 IEEE 30th Computer Security Foundations Symposium*, 2017: 276-291.

[13] He Y Y, Ni J B, Wang X Y, et al. Privacy-Preserving Partner Selection for Ride-Sharing Services[J]. *IEEE Transactions on Vehicular Technology*, 2018, 67(7): 5994-6005.

[14] Aïvodji U, Huguenin K, Huguet M, et al. SRide: A privacy-preserving ridesharing system[C]. *The 11th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, 2018: 40-50.

[15] Sherif A B T, Rabieh K, Mahmoud M M E A, et al. Privacy-Preserving Ride Sharing Scheme for Autonomous Vehicles in Big Data Era[J]. *IEEE Internet of Things Journal*, 2017, 4(2): 611-618.

[16] Li M, Zhu L H, Lin X D. Efficient and Privacy-Preserving Carpooling Using Blockchain-Assisted Vehicular Fog Computing[J]. *IEEE Internet of Things Journal*, 2019, 6(3): 4573-4584.

[17] Yu H N, Jia X H, Zhang H L, et al. PSRide: Privacy-Preserving Shared Ride Matching for Online Ride Hailing Systems[J]. *IEEE Transactions on Dependable and Secure Computing*, 2021, 18(3): 1425-1440.

[18] Fan J F, Vercauteren F. Somewhat Practical Fully Homomorphic Encryption[J]. *IACR Cryptol EPrint Arch*, 2012, 2012: 144.

[19] Paillier P. Public-Key Cryptosystems Based on Composite Degree Residuosity Classes[C]. *The 17th international conference on Theory and application of cryptographic techniques*, 1999: 223-238.

[20] Huang Y, Evans D, Katz J, et al. Faster Secure Two-Party Computation Using Garbled Circuits[C]. *The 20th USENIX conference on Security*, 2011: 35.

[21] Freedman M J, Nissim K, Pinkas B. Efficient Private Matching and Set Intersection[C]. *International Conference on the Theory and Applications of Cryptographic Techniques*, 2004: 1-19.

[22] Liu A, Zhengy K, Liz L, et al. Efficient Secure Similarity Computation on Encrypted Trajectory Data[C]. *2015 IEEE 31st International Conference on Data Engineering*, 2015: 66-77.

[23] Bringer J, Chabanne H, Favre M, et al. GSHADE: Faster Privacy-Preserving Distance Computation and Biometric Identification[C]. *The 2nd ACM workshop on Information hiding and multimedia security*, 2014: 187-198.

[24] Ishai Y, Kilian J, Nissim K, et al. Extending Oblivious Transfers Efficiently[C]. *Annual International Cryptology Conference*, 2003: 145-161.

[25] Yu H N, Jia X H, Zhang H L, et al. Efficient and Privacy-Preserving Ride Matching Using Exact Road Distance in Online Ride Hailing Services[J]. *IEEE Transactions on Services Computing*, 2022, 15(4): 1841-1854.

[26] Song D X, Wagner D, Perrig A. Practical Techniques for Searches on Encrypted Data[C]. *Proceeding 2000 IEEE Symposium on Security and Privacy*, 2002: 44-55.

[27] Hahn F, Kerschbaum F. Searchable Encryption with Secure and Efficient Updates[C]. *The 2014 ACM SIGSAC Conference on Computer and Communications Security*, 2014: 310-320.

[28] Blanton M, Steele A, Alisagari M. Data-Oblivious Graph Algorithms for Secure Computation and Outsourcing[C]. *The 8th ACM SIGSAC symposium on Information, computer and communications security*, 2013: 207-218.

[29] Mouratidis K, Yiu M L. Shortest Path Computation with no Information Leakage[J]. *Proceedings of the VLDB Endowment*, 2012, 5(8): 692-703.

[30] Carter H, Mood B, Traynor P, et al. Secure Outsourced Garbled Circuit Evaluation for Mobile Devices[J]. *Journal of Computer Security*, 2016, 24(2): 137-180.

[31] Liu C, Wang X S, Nayak K, et al. ObliVM: A Programming Framework for Secure Computation[C]. *2015 IEEE Symposium on Security and Privacy*, 2015: 359-376.

[32] Lewi K, Wu D J. Order-Revealing Encryption: New Constructions, Applications, and Lower Bounds[C]. *The 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016: 1167-1178.

[33] Das Sarma A, Gollapudi S, Najork M, et al. A Sketch-Based Distance Oracle for Web-Scale Graphs[C]. *The third ACM international conference on Web search and data mining*, 2010: 401-410.

[34] Liu C, Zhu L H, He X J, et al. Enabling Privacy-Preserving Shortest Distance Queries on Encrypted Graph Data[J]. *IEEE Transactions on Dependable and Secure Computing*, 2021, 18(1): 192-204.

[35] Akiba T, Iwata Y, Yoshida Y. Fast Exact Shortest-Path Distance Queries on Large Networks by Pruned Landmark Labeling[C]. *The*

2013 ACM SIGMOD International Conference on Management of Data, 2013: 349-360.

[36] Yu H N, Shu J G, Jia X, et al. LpRide: Lightweight and Privacy-Preserving Ride Matching over Road Networks in Online Ride Hailing Systems[J]. *IEEE Transactions on Vehicular Technology*, 2019, 68: 10418-10428.

[37] Zhang C, Zhu L H, Xu C, et al. PGAS: Privacy-Preserving Graph Encryption for Accurate Constrained Shortest Distance Queries[J]. *Information Sciences*, 2020, 506: 325-345.

[38] Liu X M, Choo K K R, Deng R H, et al. Efficient and Privacy-Preserving Outsourced Calculation of Rational Numbers[J]. *IEEE Transactions on Dependable and Secure Computing*, 2018, 15(1): 27-39.

[39] Wu D J, Zimmerman J, Planul J, et al. Privacy-Preserving Shortest Path Computation[EB/OL]. 2016: arXiv: 1601.02281. https://arxiv.org/abs/1601.02281.pdf.

[40] Coslovich L, Pesenti R, Ukovich W. A Two-Phase Insertion Technique of Unexpected Customers for a Dynamic Dial-a-Ride Problem[J]. *European Journal of Operational Research*, 2006, 175(3): 1605-1615.

[41] Ma S, Zheng Y, Wolfson O. Real-Time City-Scale Taxi Ridesharing[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2015, 27(7): 1782-1795.

[42] Yu Z Q, Liu Y, Yu X H, et al. Scalable Distributed Processing of K Nearest Neighbor Queries over Moving Objects[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2015, 27(5): 1383-1396.

[43] Gidofalvi G, Pedersen T B, Risch T, et al. Highly Scalable Trip Grouping for Large-Scale Collective Transportation Systems[C]. *The 11th international conference on Extending database technology*: *Advances in database technology*, 2008: 678-689.

[44] Ta N, Li G L, Zhao T Y, et al. An Efficient Ride-Sharing Framework for Maximizing Shared Route[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 30(2): 219-233.

[45] Santi P, Resta G, Szell M, et al. Quantifying the Benefits of Vehicle Pooling with Shareability Networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2014, 111(37): 13290-13294.

[46] Gupta A, Hajiaghayi M, Nagarajan V, et al. Dial a Ride from *k*-Forest[J]. *ACM Transactions on Algorithms*, 6(2)Article No. 41,

[47] Bei X, Zhang S. Algorithms for trip-vehicle assignment in ridesharing[C]. *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018:1-7.

[48] Huang Y, Bastani F, Jin R M, et al. Large Scale Realtime Ridesharing with Service Guarantee on Road Networks[J]. *Proceedings of the VLDB Endowment*, 2014, 7(14): 2017-2028.

[49] Tong Y, Zeng Y, Zhou Z, et al. A unified approach to route planning for shared mobility[J]. *Proceedings of the VLDB Endowment*, 2018, 11(11): 1633-1646.

[50] Ascheuer N, Krumke S O, Rambau J. Online Dial-a-Ride Problems: Minimizing the Completion Time[C]. *The 17th Annual Symposium on Theoretical Aspects of Computer Science*, 2000: 639-650.

[51] Feuerstein E, Stougie L. On-Line Single-Server Dial-a-Ride Problems[J]. *Theoretical Computer Science*, 2001, 268(1): 91-105.

[52] Xu Y, Tong Y X, Shi Y X, et al. An Efficient Insertion Operator in Dynamic Ridesharing Services[C]. *2019 IEEE 35th International Conference on Data Engineering*, 2019: 1022-1033.

[53] Waisanen H A, Shah D, Dahleh M A. A Dynamic Pickup and Delivery Problem in Mobile Networks under Information Constraints[J]. *IEEE Transactions on Automatic Control*, 2008, 53(6): 1419-1433.

[54] Alonso-Mora J, Samaranayake S, Wallar A, et al. On-Demand High-Capacity Ride-Sharing via Dynamic Trip-Vehicle Assignment[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, 114(3): 462-467.

[55] Kameswaran V, Cameron L, Dillahunt T R. Support for Social and Cultural Capital Development in Real-Time Ridesharing Services[C]. *The 2018 CHI Conference on Human Factors in Computing Systems*, 2018: 1-12.

[56] Cheng P, Xin H, Chen L. Utility-Aware Ridesharing on Road Networks[C]. *The 2017 ACM International Conference on Management of Data*, 2017: 1197-1210.

[57] Fu X, Huang J, Lu H, et al. Top-k taxi recommendation in realtime social-aware ridesharing services[C]. *International Symposium on Spatial and Temporal Databases*, 2017: 221-241.

[58] Kleiner A, Nebel B, Ziparo V A. A Mechanism for Dynamic Ride Sharing Based on Parallel Auctions[C]. *The Twenty-Second international joint conference on Artificial Intelligence - Volume Volume One*, 2011: 266-272.

[59] Santos D O, Xavier E C. Dynamic Taxi and Ridesharing: A Framework and Heuristics for the Optimization Problem[C]. *The Twenty-Third international joint conference on Artificial Intelligence*, 2013: 2885-2891.

[60] Cici B, Markopoulou A, Laoutaris N. Designing an On-Line Ride-Sharing System[C]. *The 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2015: 1-4.

[61] Vazifeh M M, Santi P, Resta G, et al. Addressing the Minimum Fleet Problem in On-Demand Urban Mobility[J]. *Nature*, 2018, 557(7706): 534-538.

[62] Asghari M, Deng D X, Shahabi C, et al. Price-Aware Real-Time Ride-Sharing at Scale: An Auction-Based Approach[C]. *The 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2016: 1-10.

[63] Zheng L, Chen L, Ye J. Order dispatch in price-aware ridesharing [J]. *The VLDB Endowment*, 2018, 11(8): 853-865.

[64] Biswas A, Gopalakrishnan R, Tulabandhula T, et al. Profit optimization in commercial ridesharing[C]. *The 16th Conference on Autonomous Agents and MultiAgent Systems*, 2017: 1481-1483.

[65] Niu B, Li Q H, Zhu X Y, et al. Achieving K-Anonymity in Privacy-Aware Location-Based Services[C]. *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, 2014: 754-762.

[66] Palanisamy B, Liu L. Attack-Resilient Mix-Zones over Road Networks: Architecture and Algorithms[J]. *IEEE Transactions on Mobile Computing*, 2015, 14(3): 495-508.

[67] Tong W, Hua J Y, Zhong S. A Jointly Differentially Private Scheduling Protocol for Ridesharing Services[J]. *IEEE Transactions on Information Forensics and Security*, 2017, 12(10): 2444-2456.

[68] Chor B, Goldreich O, Kushilevitz E, et al. Private Information Retrieval[C]. *Proceedings of IEEE 36th Annual Foundations of Computer Science*, 2002: 41-50.

[69] Zheng Y, Li M, Lou W J, et al. Location Based Handshake and Private Proximity Test with Location Tags[J]. *IEEE Transactions on Dependable and Secure Computing*, 2017, 14(4): 406-419.

[70] Ben Sasson E, Chiesa A, Garman C, et al. Zerocash: Decentralized Anonymous Payments from Bitcoin[C]. *2014 IEEE Symposium on Security and Privacy*, 2014: 459-474.

[71] Zhang H J, Deng E D, Zhu H J, et al. Smart Contract for Secure Billing in Ride-Hailing Service via Blockchain[J]. *Peer-to-Peer Networking and Applications*, 2019, 12(5): 1346-1357.

[72] Shivers R, Rahman M A, Shahriar H. Toward a Secure and Decentralized Blockchain-Based Ride-Hailing Platform for Autonomous Vehicles[EB/OL]. 2019: arXiv: 1910.00715. https://arxiv.org/abs/1910.00715.pdf.

[73] Pagnin E, Gunnarsson G, Talebi P, et al. Toppool: Time- aware optimized privacy-preserving ridesharing[J]. *Proceedings on Privacy Enhancing Technologies*, 2019, 2019(4):93-111.

[74] Aïvodji U M, Gambs S, Huguet M J, et al. Meeting Points in Ridesharing: A Privacy-Preserving Approach[J]. *Transportation Research Part C*: *Emerging Technologies*, 2016, 72: 239-253.

[75] Xu Y, Wei S Y, Wang Y S. Privacy Preserving Online Matching on Ridesharing Platforms[J]. *Neurocomputing*, 2020, 406: 371-377.

[76] Prorok A, Kumar V. Privacy-Preserving Vehicle Assignment for Mobility-on-Demand Systems[C]. *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2017: 1869-1876.

[77] Yu H N, Zhang H L, Yu X Z, et al. PGRide: Privacy-Preserving Group Ridesharing Matching in Online Ride Hailing Services[J]. *IEEE Internet of Things Journal*, 2021, 8(7): 5722-5735.

[78] Yu H N, Zhang H L, Jia X H, et al. PSafety: Privacy-Preserving Safety Monitoring in Online Ride Hailing Services[J]. *IEEE Transactions on Dependable and Secure Computing*, 2023, 20(1): 209-224.

[79] Chaudhry B, Yasar A U H, El-Amine S, et al. Passenger Safety in Ride-Sharing Services[J]. *Procedia Computer Science*, 2018, 130: 1044-1050.

[80] Wang D, Li W T, Wang P. Crytanalysis of Three Anonymous Authentication Schemes for Multi-Server Environment[J]. *Journal of Software*, 2018, 29(7): 1937-1952.
(汪定, 李文婷, 王平. 对三个多服务器环境下匿名认证协议的分析[J]. *软件学报*, 2018, 29(7): 1937-1952.)

[81] Wang Z, Fan J, Cheng L, et al. Supervised Anonymous Authentication Scheme[J]. *Journal of Software*, 2019, 30(6): 1705-1720.
(王震, 范佳, 成林, 等. 可监管匿名认证方案[J]. *软件学报*, 2019, 30(6): 1705-1720.)

[82] Miers I, Garman C, Green M, et al. Zerocoin: Anonymous Distributed E-Cash from Bitcoin[C]. *2013 IEEE Symposium on Security and Privacy*, 2013: 397-411.

[83] Kumar A, Fischer C, Tople S, et al. A traceability analysis of monero's blockchain[C]. *European Symposium on Research in Computer Security*, 2017: 153-173.

于海宁 于 2013 年在哈尔滨工业大学信息安全专业获得博士学位。现任哈尔滨工业大学网络空间安全学院副研究员。研究领域为数据安全、隐私计算等。Email: yuhaining@hit.edu.cn

张宏莉 于 1999 年在哈尔滨工业大学计算机系统结构专业获得博士学位。哈尔滨工业大学网络空间安全学院教授。研究领域为网络与信息安全、网络测量与建模、网络计算、并行处理等。Email: zhanghongli@hit.edu.cn

余翔湛 于 2005 年在哈尔滨工业大学计算机系统结构专业获得博士学位。哈尔滨工业大学网络空间安全学院教授。研究领域为信息安全、网络流量分析等。Email: yxz@hit.edu.cn

曲家兴 于 2019 年在哈尔滨工程大学计算机应用技术专业获得博士学位。黑龙江省网络空间研究中心教授级高级工程师。研究领域为网络安全、网络舆情分析等。Email: smilingqu@126.com