文章编号:1671-4598(2020)10-0200-05 DOI:10.16526/j.cnki.11—4762/tp.2020.10.041 中图分类号:TP311 文献标识码:A

车联网下大数据安全采集机制研究

赵 明,王英资
(中汽数据(天津)有限公司,天津 300300)

摘要:现存的物联网协议不能够直接应用于大数据采集场景,并且常规动态网络结构和车辆节点的复杂性会随着数据量增多而加大,对安全性要求变得越来越高;针对车辆数据资源传输的安全性保证问题,提出了一个新的大数据收集安全机制;车辆信息通过大数据注册中心连接到网络后进行联合,以“身份验证”和“单点登录”的方式实现大数据中心算法;提出新的安全的大数据收集方案,处理大型车辆数据结构与汇聚节点数据,进行二次登录再到数据安全收集存储,完成安全性能评估;研究结果表明:新的数据安全交换算法使消息摘要和随机密钥开销降低,车联网中大数据的效率提升了21.67%,车辆节点方式可以提升26.41%。

关键词:车联网;大数据;采集;安全机制;物联网;安全

Research on Big Data Security Collection Mechanism under the Internet of Vehicles

Zhao Ming, Wang Yingzi
(Automotive Data of China (Tianjin) Co., Ltd., Tianjin 300300, China)

Abstract: Existing IoT protocols cannot be directly applied to big data collection scenarios, and the complexity of conventional dynamic network structures and vehicle nodes will increase as the amount of data increases, and safety requirements become higher and higher. Aiming at the security guarantee of vehicle data resource transmission, a new big data collection safety mechanism is proposed. After the vehicle information is connected to the network through the big data registration center, the information is combined, and the big data center algorithm is implemented by " identification" and " single sign—on". A new secure big data collection scheme is proposed, processing large—scale vehicle data structure and aggregation node data, performing a second login and then to data security collection and storage to complete the safety performance evaluation. The research results show that the new data security exchange algorithm reduces the message digest and random key overhead, increases the efficiency of big data in the Internet of Vehicles by 21.67%, and increases the vehicle node mode by 26.41%.

Keywords: internet of vehicle; big data; collection; secure mechanism; Internet of things; security

0 引言

车联网(IoV, Internet of Vehicles),作为物联网(IoT, internet of things)的一个重要分支,是实现智能化的统一管理的有效途径,是智慧城市的重要标志,具有广泛的应用前景[1]。随着IoV的发展,越来越多的车连接到互联网,在网上传播和共享数据。大规模的IoV从不同的地方收集数据,这些数据不论在大小,还是数量上都符合大数据的异构性和维度。

在IoV中,车辆的轨迹使得IOV能够通过有线或无线的渠道分享交通信息,为交通智能管理和路径优化起到较好的作用[2]。随着社会的发展,越来越多的车辆和道路使得IoV规模扩大。车辆连接网络,并安装了各种传感器,包括温度、加速度等,能够获取车辆的相关信息,包括驾驶状态和交通状况等[3]。由于其依赖性,数据在时间上具有时空性。

随着IoV的传播和发展,收集的各种交通信息会涉及到个人隐私,例如所监控范围内车辆的实时运行位置情况,甚至还包括相关车辆运行参数以及交通安全重要数据[4]。但由于车辆节点的存在,可能会发送一些欺诈、假冒信息来危害交通,造成不必要的麻烦[5]。因此,设计一个确保车辆数据的传输资源是可信任的,不会被篡改是一个很重要的安全机制。智能交通系统不断发展和应用于IoV[6],车辆和车辆之间的大数据收集应用平台变得越来越频繁,如何实现大数据收集的安全机制是有意义的,值得研究的课题。

很多研究者对大数据和IoV安全性开展了广泛的研究。文献[7]等提出了一种在用户和RSU之间交换数据消息的安全方案,但IoV的可扩展性仍然是一个待解决的问题。文献[5]提出了一种平衡公共安全和车辆隐私的有效系统,以保证信息的可信赖性。文献[8]提出了一种安全机制,用于保护与车辆到网格(V2G)网络中的可用加密原语的保密通信。文献[9]在大数据领域工作并开发了安全和隐私机制。作为大数据领域的一项重要技术,Hadoop的安全性得到了解决[10]。文献[11]提出了一种用于大数据

收稿日期:2020-03-22; **修回日期**:2020-04-07。

作者简介:赵 明(1986-),男,天津人,硕士,工程师,主要从事模型计算、诊断算法、故障分析方向的研究。

安全调度的密钥交换方案。文献［12］中也提出了解决相关领域认证和隐私问题的安全模型。文献［13］使用“利弊解决方案距离法”确定和改善 IOV 信息服务提供商的决策者与消费者之间的差距，以此来改善车辆远程信息处理系统，对于该系统的功能进行完善。文献［14］使用计算机电子表格程序来开发新模型，该模型可以使关键交通参数更灵活地进入数据库，以帮助评估车辆高速公路系统的影响。Thomas F. Golob 分析了互联车辆信息服务系统中的信息技术，发现了车联网在个人旅游信息的采集方面，由于数据量的关系，仍然不能大规模应用[15-16]。但是现有的协议还不能够直接应用于本文的大规模 IoV 的大数据采集场景。因此，大数据收集的安全性和效率问题仍然值得研究。

本文通过提出一个信息收集方案，运用方案中的“单点登录算法”提升登录效率，在车辆节点、数据节点、大数据之间建立节点关系和特有的数据结构以实现数据安全存储；运用提出的“证书”对签名节点进行身份验证来避免的消息被恶意攻击，从而提升消息处理的安全性；最后通过数据模拟实验验证提出的安全采集机制，以显著提升信息收集的安全性和高效性。

1　方案设计

1.1　车联网（IoV）架构

车联网（IoV）是按照特定的数据交互标准和通信协议，并基于车载 Ad Hoc 网络和车载移动互联网集成的网络车载网络。这是一个扩展的应用程序，能够实现车辆智能化控制和智能动态信息服务[17]。

大数据中心、车辆节点、汇聚节点这三类构成车联网的基本架构。大数据中心主要功能是对车辆节点收集来的交通数据进行管理和处理。在车辆节点的车载单元内，车辆网关是从定位模块和获取模块那里收集数据。在汇聚节点阶段，主要通过协同用户和路边单元相互之间的信息通信，以转移相关信息，如图 1 所示。

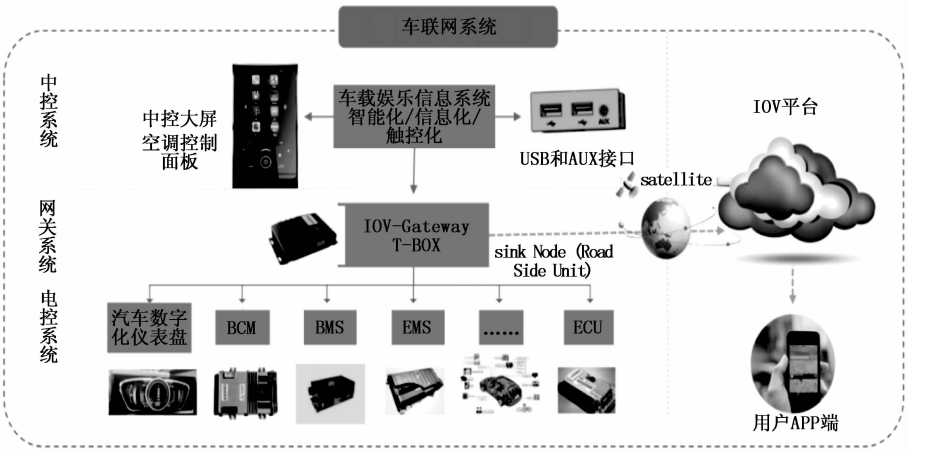

图 1　车联网基本架构

与传统的 Ad hoc 网络相比较，IoV 具有很多不同的功能。由于车辆节点能高速改变其所在位置，因此节点拓扑结构是动态的并且在变化。想要建立准确的社区是很困难的。

1.2　系统模型设计

为了满足大规模 IoV 的安全要求，本文提出了一种用于大数据的安全数据收集方案。越来越多的大型车辆节点从不同的地方生成各种属性数据。这些数据将由具有安全保护的大数据中心收集，并使用 Hadoop 架构存储在分布式存储系统中。

大数据是一个让大量数字化并将其与现有数据库相结合的系统信息。它是基于 3 个主要特征定义的，也称为 3V：高容量、高速度和多类型[18]。越来越多的车辆从不同地方收集数据，汇集了异构的大数据。该大数据和 IoV 的整合已成为一种趋势，推动了新的信息技术的发展[19]。大数据收集可以改善决策，尤其是在 IoV 中进行路径规划。对于政府，收集的大数据能够帮助分析和解决交通问题。对于像实时运输这样的公司，它帮助优化车辆资源。因而，政府和公司要求并开始构建车联大数据平台。

在初始化阶段，与所有新增加的车辆节点的认证相关联形成了针对非法节点的第一防御安全线。这些节点将在系统中注册，并与数据中心交换必要的信息。系统在完成初始化阶段后，运用之前提出的安全单点登录算法，使得登录协议的效率提升了。与此同时收集到的信息在相对处于“安全保护”的情况下进行传递，直到该节点注销。表 1 为本文使用的参数定义及表述。

表 1　参数定义

参数	定义
V	车辆节点
S	汇聚节点
C	大数据中心
r	随机数
Ts	时间戳
r	随机数
Cert_veh	车辆节点证书
Cert_cen	大数据中心证书
sign_sink	汇聚节点签名
pk_sink	汇聚节点公钥
key_sc	会话密钥(SC)
key_vc	会话密钥(VC)
M1	业务数据
m1,m2,m3,m4	参数
$HMAC$/m	MAC 计算
H/m	Hash 值

1.3　模型初始化

为了支持不同类型的大数据平台，假设每辆车都配备

了由外部认证机构（CA）颁发的证书。在初始化阶段，每个车辆都需要在大数据中心进行注册后方能接入网络。这是一个必要环节，注册后在大数据中心与车辆节点内部分别生成各自的公钥和私钥。如图 2 所示，大数据中心和车辆节点完成相应的公钥交换认证。如果证书通过检查，相应的 ID 将被注册为有效帐户。接收节点负责消息转发，其中，汇聚节点也是在这个阶段注册的。

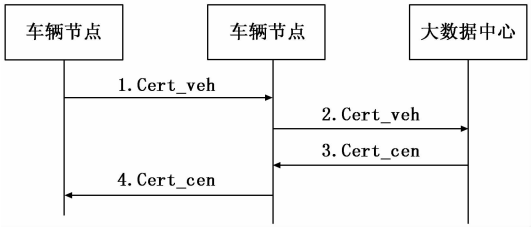

图 2 初始化阶段消息交换

1.4 首次登录

随着 IOV 日新月异的发展，使得更多的车辆连接到网络中。车辆可以在高速行驶的过程中，将信息同步连接到不同的汇聚节点。利用车联网架构中的高动态拓扑结构，节点必须通过单点登录，以此来实现授权，保证只有授权的节点才能对资源进行访问。在信息收集方案中提出了的“单点登录算法”，该算法的实施使得登录协议的运行效率得以提高。利用所提出的方案增强了可扩展性。在初始化阶段完成之后，利用不同的协议，汇聚节点和车辆节点分别连接到大数据中心，如图 3 所示。

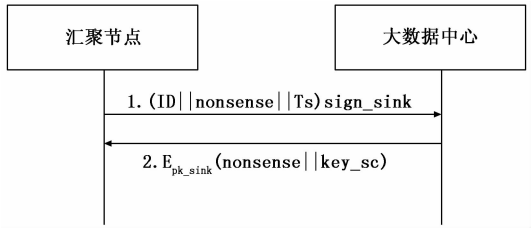

图 3 首次登录

在汇聚节点登录阶段，ID，r 和 Ts 被发送到具有汇聚节点签名的大数据中心，如图 3 所示。根据收到的消息，大数据中心检查汇聚节点的签名和 ID。Ts 保证了时间效率，能够抵御重播攻击。如果消息来自于有效合法的帐户，则大数据中心会生成唯一的 key _ sc。r 和 key _ sc 将使用 pk _ cen 加密，然后发送到汇聚节点。在使用 sk _ sink 解密密文后，汇聚节点将获取 key _ sc。

1.5 再次登录

当车辆节点离开其第一个登录汇聚节点的区域时，它必须通过另一个登录访问新到达的汇聚节点。对于这种车辆节点，所提出的方案仅仅是之后的登录过程。如图 4 所示，车辆节点和汇聚节点之间的交互可以提高登录过程的效率并更新会话密钥 除了存储的“票证”m2 之外，证书和 Ts 被发送到具有车辆节点签名的汇聚节点。m2 中的大数据中心签名证明该票由大数据中心授予。如果证书中的 ID 与 m2 中的 ID 匹配且时间戳未超过期限，则该车辆将被视为合法节点并登录系统。由 pk _ sink 加密的汇聚节点证书和 key _ vs 随后将被发送到车辆节点。

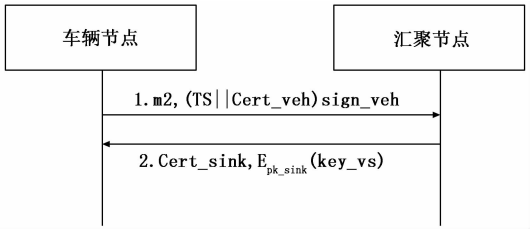

图 4 车辆节点再次登录

1.6 安全数据收集

本节为大规模车载互联网中的数据收集设定了安全前提。在车辆节点成功登录系统的情况下，将使用以下算法收集业务数据，如图 5 所示。

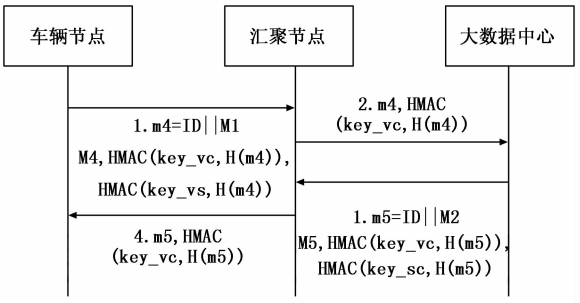

图 5 业务数据收集的消息交换

M1 和 M2 表示作为主要交互对象的业务数据。温度参数等业务数据可以纯文本形式传输。m4 通过车辆节点 ID 和 M1 的串联来计算。为了提高计算效率，利用 m4 的哈希值来计算。由于 key _ vc 和 key _ vs 已预先共享，因此 HMAC 有助于防止篡改数据并保证数据发送方的身份。汇聚节点验证 HMAC（key _ vs，H（m4））并发送 HMAC（key _ vc，H（m4））。当大数据中心发布 M2 时，在步骤 3 和 4 中提出相同的算法。

1.7 安全数据存储

在上面的小节中，提出了一种用于大数据的安全信息收集方案。系统将车辆和汇聚节点的必要的安全数据存储在大数据中心里面。其中，“车辆”与“汇聚节点”所包含的一些必要的安全数据在“大数据中心”存储，存储在大数据中心的“车辆数据结构”与“汇聚节点数据结构”的设计，如表 2 所示。表 2 中“ID”和“证书”分别表示相关的节点标识。一旦车辆的对应节点在系统中成功的注册后，则系统相关状态会从“关闭”变成“开启”。如果通过大数据中心的分析检测到节点有异常动作，则它的状态将由“开启”变为“关闭”。如果时间戳超出有效期，则节点必须添加新的身份登录。

表 2　车辆和汇集节点的数据结构

节点	ID	证书	状态	有效阶段	密钥
车辆节点	ID_veh	Cert_veh	On/off	Ts_Period	Key_vc
汇聚节点	ID_sink	Cert_sink	On/off	Ts_Period	Key_sc

在数据存储算法中，本文所建立的算法能高效的实现授权，以确保把通过授权的节点作为访问资源的唯一通道，保证安全性。

2　安全性能评估

2.1　安全分析

随着先进科学信息技术的不断发展，大规模的 IoV 具有巨大的商业利益和研究价值。大数据采集的安全性具有重要意义。为确保数据收集的安全性，并满足认证，完整性，机密性，不可否认性和授权等要求。

根据 IoV 的特点，提供安全信息收集方案必须符合要求，以确保数据收集的安全性。需要包括 5 种安全需求：

1）通过认证的方式得以识别车辆的相关节点，汇聚节点和大数据中心；

2）满足相关资料和信息数据的完整性要求，保护消息不被修改或破坏；

3）通过授权的方式保证通过授权节点一种途径来访问资源。在高动态拓扑结构中，节点须拥有属于自身的单点登录机制。

4）不可否认性以防止后来否认；

5）机密性，即保护发送信息的实体。诸如“温度”等公开数据，它们的参数可以通过纯文本形式传输，而“传输位置数据”等机密数据则必须以“密文”的形式进行传递；

除了安全要求之外，所提出的机制还阻止了诸如中间人（MITM）攻击，重放攻击，伪装攻击和消息操纵攻击等安全攻击。使用 CA 颁发的证书对系统中的节点进行身份验证。同传统的“用户名＋密码令牌”方案相比，运用所提出方案中的“证书”可以抵抗暴力性破解，而且不能伪造，这对于认证更可靠。从“初始化阶段”开始到“数据收集阶段”所进行的信息交换中，使用签名来确保完整性以防止修改或破坏。在初始化阶段交换公钥，私钥帮助加密。结合公钥/私钥，对称密钥保护要发送给适当实体的信息。为了满足保密要求，机密数据以密文形式传输。私钥存储，私钥用于计算不可否认性的签名，以防止否认发送过该条信息。在数据存储器中，会话密钥“key _ vc”是作用到控制大数据应用程序中，从而访问相关车辆节点的数据。只有拥有了证书对应的授权节点，才能有效访问该资源。

在方案中，m2 被用作单点登录的“票证”。但是，攻击者无法成功登录，即使它复制了之前的 m2，证书和 Ts 也需要检查车辆节点的身份。因此，恶意节点无法根据提出的机制对大数据中心进行重放攻击。在提议的机制中传输的消息使用会话密钥加密，签名对于 MITM 的攻击有一定的对抗和防御作用。在伪装攻击中，攻击者伪装成有效节点来发送错误消息对信息系统安全性有不良影响。在所提出的方案中，使用证书和签名对大规模网络中的所有节点进行认证。因此，伪装攻击者无法在有效节点之间发送错误消息，因为伪装攻击者是无法通过身份验证并伪装成有效节点。在消息处理攻击中，交换的消息可能被丢弃，修改甚至伪造以中断攻击者的数据收集。此机制使攻击者难以伪造数据包或路径。因此，使用此机制进行消息操纵攻击无效。

2.2　总体实验时间

为了评估所提出的安全机制的性能，使用网络模拟器软件 Opnet 进行整个数据收集过程的模拟[20]。如图 6 所示，给出了 3 种节点的总时间信息。提供了汇聚节点，大数据中心和车辆节点之间完整安全的大数据收集，以显示工作流程的进展情况。正如 IEEE 802.11 中关于车辆到基础设施（V2I）中信息和电信通过相互交换所定义的那样，假使本文应用场景的传输速率最高可达每秒 12 兆字节，横坐标轴标识系统的交互时间，纵轴表示在运用本文的安全协议后节点的传输速率。采取连续模拟交互的方式，主要包括 4 个步骤：1）初始化阶段；2）汇聚节点登录；3）车辆节点登录；4）数据收集。对于图 6 中单个节点，当蓝线超过零的时间段长度，则表示每一个动作传输的时间。两个交互对象计算时间则由两个后续动作之间的差值时间值进行标识。实验结果表明，该机制可用于大规模 IoV 环境。

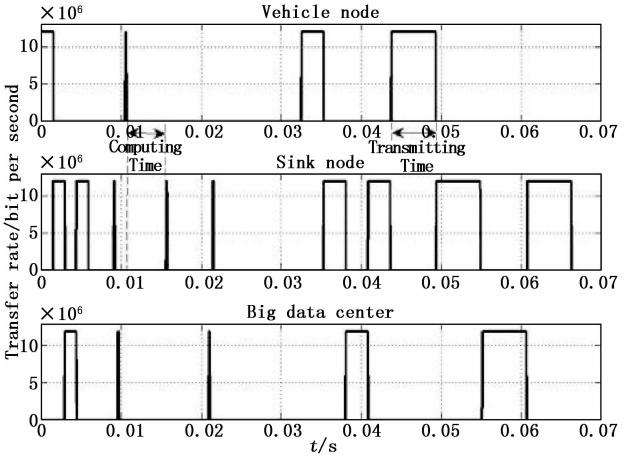

图 6　方案的总体时间

2.3　计算时间

在方案中，消息摘要和随机密钥（Tk）用于提高效率。使用 HMAC 算法直接处理消息而不在方案 1 和 3 中预先计算消息摘要，同样地，数据在方案 2 中直接使用

key _ vc 和 key _ vs 的会话密钥加密。经过多次试验如图 7 所示，如表 3 所示使用大数据中心方案的计算时间比使用其他方案的计算时间短平均缩短 21.67%，使用车辆节点的计算时间比使用其他方案的计算时间缩短 26.41%。使用汇聚节点方案的计算时间几乎与使用其他方案的计算时间是近乎相同的。

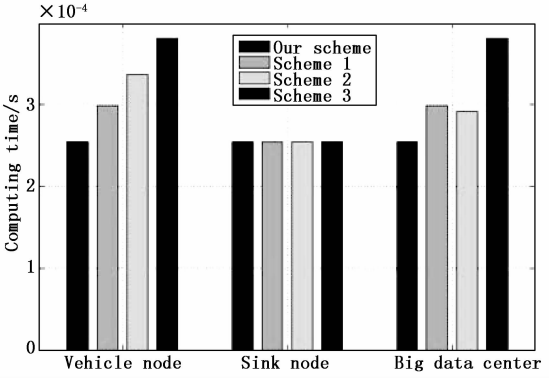

图 7 计算时间比较

表 3 数据安全采集相比其他方案开销时间统计

名称	时间(单位:10^{-4})			
	Our scheme	Scheme 1	Scheme 2	Scheme 3
Vehicle node	2.51	2.99	3.42	3.82
Sink node	2.51	2.50	2.48	2.49
Big data center	2.53	2.99	2.91	3.79

3 结束语

本文提出了一种大规模 IoV 大数据的安全信息收集方案，用于认证的单点登录算法，通过登录系统运行使得系统计算开销降低效率得以提升。运用节点机制在汇聚节点、车辆节点、大数据中心三者之间应用实现信息传递，进行有效算法授权提升登录安全性。提出的安全数据交换算法运用密钥授权方式防止了常用的安全攻击。

同时运用 OPNET 网络仿真应用软件对建立的机制性能进行数据收集，通过对汇聚节点、大数据中心以及车辆节点进行大数据收集并对收集内容进行量化分析，在一定交互时间内统计“交互速率”，根据数据分析验证了该机制适用性。通过对比数据结果得出应用于机制后的大数据中心、车辆节点比常规形式要节省时间，大数据中心缩短 21.67%，车辆节点缩短 26.41%，从而使用消息摘要和随机密钥有助于减少开销，车联网中大数据的运行效率得以提升。

综上所述，本文提出的安全信息收集方案在大规模 IoV 中的大数据应用中可以提高效率，具有安全性和可落地性。

参考文献：

[1] 程 刚，郭 达. 车联网现状与发展研究 [J]. 移动通信，2011，35 (17)：23 - 26.

[2] 孙宝林，李腊元. 多跳无线移动 AdHoc 网络路由协议的研究分析 [J]. 小型微型计算机系统，2004，25 (10)：1737 - 1741.

[3] Li B，Zhao C，Zhang H，et al. Characterization on clustered propagations of UWB sensors in vehicle Cabin：measurement，modeling and evaluation [J]. IEEE Sensors Journal，2013，13 (4)：1288 - 1300.

[4] Zhou Y，Chen S，Zhou Y，et al. Privacy－preserving multi－point traffic volume measurement through vehicle－to－infrastructure communications [J]. IEEE Transactions on Vehicular Technology，2015，64 (12)：5619 - 5630.

[5] 马得途. 基于 LTE－R 的车车通信技术及可靠性研究 [D]. 兰州：兰州交通大学，2018.

[6] Soares J，Borges N，Canizes B，et al. Probabilistic estimation of the state of electric vehicles for smart grid applications in big data context [A]. 2015 IEEE Power & Energy Society General Meeting [C]. IEEE，2015.

[7] 柯敏毅，魏树婧，肖 鹏. 移动 AdHoc 网络路由安全问题研究 [J]. 网络安全技术与应用，2008 (2)：41 - 43.

[8] Wang H，Bo Q，Wu Q，et al. TPP：traceable privacy－preserving communication and precise reward for vehicle－to－grid networks in smart grids [J]. IEEE Transactions on Information Forensics & Security，2017，10 (11)：2340 - 2351.

[9] 杨炜圻. 大数据环境下的数据安全性探讨 [J]. 网络安全技术与应用，2014 (8)：136 - 136.

[10] 孙珊珊. 基于 Hadoop 的云计算数据安全性研究 [D]. 武汉：武汉理工大学，2013.

[11] 曾 怡. 大数据应用在云计算平台的优化部署与调度策略研究 [J]. 海峡科技与产业，2017 (9)：87 - 88.

[12] Li H，Lu R，Zhou L，et al. An Efficient Merkle－tree－based authentication scheme for smart grid [J]. IEEE Systems Journal，2014，8 (2)：655 - 663.

[13] 李军侠. 基于三步搜索算法的解距离模糊方法 [J]. 系统工程与电子技术，2011，33 (3)：557 - 561.

[14] 李启彬，蒋国斌，徐令伦，等. 影响高速公路交通噪声预测关键参数 [J]. 噪声与振动控制，2015 (2)：126 - 129.

[15] 郭 鑫. 旅游大数据与挖掘分析研究 [J]. 电脑知识与技术，2013 (14)：3215 - 3216.

[16] 金雯婷，张 松. 互联网大数据采集与处理的关键技术研究 [J]. 中国金融电脑，2014 (11)：70 - 73.

[17] Alam K M，Saini M，Saddik A E. Toward social internet of vehicles：concept，architecture，and applications [J]. IEEE Access，2015，3：343 - 357.

[18] 孟小峰，慈 祥. 大数据管理：概念、技术与挑战 [J]. 计算机研究与发展，2013，50 (1)：146 - 169.

[19] 冯登国，张 敏，李 昊. 大数据安全与隐私保护 [J]. 计算机学报，2014，37 (1)：246 - 258.

[20] Guo J H，Xiang W D，Wang S Q. Reinforce networking theory with OPNET simulation [J]. Journal of Information Technology Education，2007，6 (6)：215 - 226.