IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON DISTRIBUTED COMPUTING INFRASTRUCTURE FOR CYBER-PHYSICAL SYSTEMS

Received May 4, 2019, accepted May 25, 2019, date of publication June 7, 2019, date of current version July 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2921605

Pseudonym Management Through Blockchain: Cost-Efficient Privacy Preservation on Intelligent Transportation Systems

SHIHAN BAO[1], YUE CAO[2], AO LEI[3], PHILIP ASUQUO[1], HAITHAM CRUICKSHANK[1], ZHILI SUN[1], AND MICHAEL HUTH[4]

[1]Institute of Communication Systems, University of Surrey, Surrey GU2 7XH, U.K.
[2]School of Transportation Science and Engineering, Beihang University, Beijing 102200, China
[3]Huawei Technologies, Beijing 100085, China
[4]Department of Computing, Imperial College London, London SW7 2AZ, U.K.

Corresponding author: Yue Cao (yue.cao@lancaster.ac.uk)

This work was supported in part by The PETRAS Internet of Things Research Hub. The authors would like to thank Prof. C. Maple for the support.

ABSTRACT Research into the established area of the intelligent transportation system is evolving into the Internet of Vehicles, a fast-moving research area, fuelled in part by rapid changes based on cyber-physical systems. It needs to be recognized that existing vehicular communication systems are susceptible to privacy vulnerabilities which require addressing. A practical challenge is that many vehicular communication applications and services make use of basic safety messages that contain the identity of the vehicle, location, and other personal data. A popular way of dealing with this privacy issue is to utilize a pseudonym change scheme to protect the vehicle's identity and location. However, many such schemes suffer that the cost grows and the certificate management difficulty raises with the number of pseudonyms generated and stored, casting doubt of the economic feasibility of that approach. We propose a decentralized blockchain-based solution for pseudonym management that overcomes these limitations. This scheme consists of pseudonym distribution and a shuffle operation, allowing the reuse of existing pseudonyms to different vehicles. The results reported here, including those from our simulations, demonstrate that the proposed scheme can reuse existing pseudonyms and achieve a better degree of anonymity at a lower cost than existing schemes.

INDEX TERMS Pseudonym shuffling, blockchain, transportation-based cyber-physical systems, vehicular communication system.

I. INTRODUCTION

Cyber-Physical system (CPS) could be considered as one of the most promising techniques to help people live a better life. One of the most attractive CPS cases is the Intelligent Transportation Systems (ITS), as denoted as the Transportation-based Cyber-Physical System (TCPS). The combination of vehicle and network communication technologies has pushed the boundary of next generation, connected vehicles. This exerts pressure on car manufacturers to offer innovative products and services in that space. While the connected vehicle and roadside infrastructure are physical entities, the Vehicular Communication System (VCS) is a network platform that provides Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communications. With the help of the development of distributed computing infrastructures for CPS, the vehicle becomes a platform capable of receiving information from its peers and the environment, generating its own data, such as driver behavior and car state, and transmitting data to other vehicles, roadside infrastructure, or third parties in order to improve road safety, pollution control, insurance information and traffic efficiency.

In addition, the Internet of Things (IoT) technology is driving traditional VCS research and development towards the Internet of Vehicles (IoV) [1]. Applications in IoV rely on the exchange of Basic Safety Messages (BSMs) which contain vehicle status information such as location, speed, and vehicle dimension [2]. Due to the fact that many applications and services make use of BSM – which contains vehicle identity, location and other personal data – VCS faces the risk of

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Yu.

not only disclosing sensitive information about vehicles and users, but also of adversarial manipulation of identity and location information.

Existing pseudonym and certificate management systems still left few challenges to overcome. A common solution called Security Credential Management System (SCMS) [3] has been well investigated, by providing a large scale system which can support 300 billion certificates per year for 300 million devices at full capacity. However, this advantage comes with a shortcoming: the system would have large certificate revocation lists and would be difficult and inefficient to achieve certificate revocation. The authors in [3] themselves illustrate the method of SCMS is prohibitively expensive regarding to storage limitations on the device (OnBoard Unit). This constitutes the motivation for the research we develop and report in this paper.

In a traditional VCS structure, a central manager such as a Certificate Authority (CA) or Public Key Infrastructure (PKI) is designed to manage pseudonyms certificate *centrally*. However, a centralized network can be highly unstable, have low scalability, and represents a significant single point of attack. A number of pseudonym management schemes state that a distributed and decentralized system could achieve better anonymity and durability [4], [5]. Since different locations would have different demands on pseudonym availability – based on the traffic and other factors – the assignment of pseudonyms is challenged by the variability of such needs. However until now opinion suggests that decentralized RSUs appear to be unable to handle the pseudonym assignment problem efficiently – e.g. the paper [4] proposes a roadside unit (RSU) assisting pseudonym reused scheme using a distributed optimization algorithm. Although the paper mention the distributed optimization algorithm, there is no fully explanation about how they fit that in their system.

With all this in mind, we posit that blockchain technology and distributed ledgers [6] could be a feasible tool for resolving the challenges above. To tackle distribution optimization problem in the shuffling process without a central manager, the pseudonym shuffling is realized by using the Blockchain distributed consensus. The pseudonym shuffling results are recorded in blocks (distributed ledger). The method also provides randomness of pseudonym shuffling and fully traceable record for certification revocation use. The blockchain technology brings robustness in the distributed structure. When a single point fails, the rest would still continue to work. The method also provides randomness of pseudonym shuffling and fully traceable record for certification revocation use. The details of pseudonym shuffling is available in section III-C.2.

We propose a framework for providing privacy-preserving pseudonym management that is more cost-effective across the system lifecycle than existing approaches. Firstly, a pseudonym Management scheme by using blockchain technology is proposed as the first contribution. Secondly, we introduce pseudonym certificate shuffling scheme, which is a new location privacy preservation scheme for VCS. It reduces pseudonym generation and management cost.

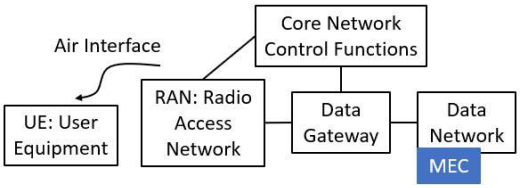

FIGURE 1. A Brief 5G architecture.

A decentralized privacy manager (PM) is introduced in the system afterwards. The PM aims to alleviate the computation burden on RSUs and to improve the robustness of the network. As shown in **Fig.1**, PM can be deployed as Multi-access Edge Computing (MEC) node within data network. The data gateway forwards data from MEC to the Radio Access Network (RAN, e.g. 5G base stations) and User Equipment (UE, e.g. mobile phones) access the MEC via the air interface between UE and RAN. Finally, asymmetric cryptography is used in blockchain transactions to protect pseudonym shuffle path. Each transaction is signed with sending PM's private key and encrypted by receiving PM's public key. As a result, either other PMs in the blockchain network or attackers outside of the network cannot observe the information from this specific transaction.

The remainder of this paper is organized as follows: Section II briefly introduces related techniques. The model overview and details of our scheme are discussed in section III; we describe our system model, including the shuffling algorithms. The scenario for attack analysis and performance evaluation is given in Section IV. Section V concludes the paper and presents some plans for future work.

II. RELATED WORK

In this section, we review the characteristics of related schemes and then offer a brief literature review about extant work on privacy preservation in the IoV.

A. PRIVACY-PRESERVING SCHEMES

In recent years, technologically-realized preservation of privacy has attracted a lot of attention in the research community. One of the most widely acceptable solutions for preserving privacy in mobile environments is the use of *pseudonyms*. These are temporary identifiers of vehicles instead of a fixed real identity of a vehicle.

It is important for privacy that the original identity of a vehicle is never used to sign vehicular network messages. However, this original identity may serve as input for generating or requesting pseudonyms from a Certificate Authority (CA). Such pseudonyms and related certificates are only valid if also signed by a CA. A vehicle holds a set of pseudonyms it can store locally and use as temporary addresses for signing and sending messages over a wireless channel.

A common method to prevent linking different pseudonyms to the same vehicle (and so revealing the identity of the vehicle) is to change the pseudonym of vehicles based on a time or action domain. This is problematic as

privacy demands particular frequencies of such pseudonym change [7]. There are numerous proposals for pseudonym change schemes in VANETs. However, there is no common agreement on the most effective scheme or the most suitable solution strategy. The authors in [8] and [9] propose simple approaches for exchanging pseudonyms in a fixed or random time, namely coordinated silent period (CSP) and cooperative pseudonym change (PCN). However, the required basic safety message is still linkable in these approaches, be it through matching similar pseudonyms or by reconstructing vehicle traces from the broadcast messages [10], [11]. A potential solution was proposed by the paper Security Credential Management System (SCMS) [3] which claims one of the leading PKI candidate designs in the United States. One advantage about the design is that the system could support 300 billion certificates per year for 300 million devices at full capacity. In SCMS, the paper states that the lifetime of each certificate is specific 5 minutes and one vehicle would carry up to 3 years' worth of pseudonym certificates which are more than 300, 000 certificates. The authors in [3] illustrate this method is prohibitively expensive regarding to storage limitations on the device (On-Board Unit). Moreover, the large amount of certificates would significantly increase the size of certificate revocation list (CRL), which reduces the efficiency in terms of pseudonym certificate revocation and takes up the bandwidth usage. Therefore, having a sustainable life-cycle of pseudonym certificate is crucial.

B. BLOCKCHAIN AND BLOCKCHAIN-BASED APPLICATIONS

Nowadays, Bitcoin attracts a lot of attention along with its blockchain concept, which was proposed in 2008 [6]. In simple terms, a blockchain is a synchronized and distributed ledger which stores a list of blocks. Each block records a set of validated transactions (e.g. user information and a receipt) and securely links to the previous block. Central managers are removed from the blockchain structure and the public ledger is maintained by all the network participants instead. This is realized by a protocol that achieves a trustworthy consensus about the chain of blocks created. In other words, network nodes can agree (deterministically) on the history and order of blocks that were created, and on which node is allowed to add the next block to the chain.

The leader election of the node that can add the next block may be performed through a variety of techniques. For example, Proof of Work poses a cryptographic puzzle to nodes based on a cryptographic hash function, the last local block seen, and the pool of transactions to be processed at a local node. A node that solves this puzzle announces the solution on the network, and other nodes accept such solution only if all transactions in the new block validate, the block does correctly point to the last block, and no other such solution was received beforehand. Since solving a puzzle is hard but verifying a solution is easy, this system provides security and effective validation and does not have a single point of failure.

The network will reach *eventual consistency* since some regions may temporarily diverge in their opinion of who won the next block. Since nodes hold an entire block *tree*, such disputes get resolved eventually as all nodes consider the path in the local tree with the ''biggest overall work'' to be the genuine chain (and this choice may vary over time).

Blockchain offers a means of creating a trustworthy record of transaction histories in a network of nodes in which there exists mistrust. This is a conservative trust model for VCS, where some parts of the network would be within trusted computing bases (e.g. the CA) but other parts would be more open or even publicly accessible (e.g. the vehicles as nodes).

Blockchain security is achieved in a manner that reflects the design choices of the blockchain. For example, when Proof of Work is used for consensus, then one would need to control more than 50% of the nodes in the network in order to rewrite the blockchain history and so corrupt data veracity [12]. This high degree of resiliency is what makes blockchains attractive in settings in which faults and malicious manipulation may corrupt integrity of data ledgers.

Blockchains are beginning to be used not only for decentralized cryptocurrencies, but also for a wide range of applications including those in Internet-of-Things (IoT) scenarios [13]–[15], and [16].

Despite the fact that blockchain has received a lot of attention from the banking industry, people find that the use of blockchain can also improve other systems such as insurance, electric vehicles charging and car sharing services [15]. The paper [12] states that there are some concerns about Blockchain, namely, majority attack, selfish mining, identity disclosure and abuse of Blockchain.

C. PRIVACY ATTACKS IN IOV

In [17], we published a survey that comprehensively analyzed security and privacy requirements in vehicular networks. Privacy threats were studied and classified into the following categories of attack. The Trace Analysis attack is used for tracking a mobile phone. The historical cloaking regions are linked to the mobility pattern of the user. A location-based system (LBS) server can derive probabilities of the mobile user being at different locations of the cloaked region [18]. Bogus location proofs are generated when two nodes collude with each other. For example, if a malicious node m_1 needs to assert that it is in a location at which it is not, it can have another colluding node m_2 to mutually generate bogus location assertions for it [19].

The authors in [20] present Trajectory Attacks as a location privacy attack where an adversary uses the knowledge of the user's locations to link the user location to a particular query. Trajectory attacks are possible even if the identifier of the user has been removed [21].

Attacks on data integrity will fail to provide the trusted services to the users and vehicles. The authors in [22] have evaluated the attacks on data integrity on real-time traffic information manipulation that is generated and passed by the vehicles in the ITS. The paper states that data integrity attacks

could disrupt the ITS service information and even cause a severe traffic congestion.

Lastly, the Transition Attack is one in which the adversary uses previous observations to estimate the transition probability for each possible turn at intersections [23]. The adversary tries to reconstruct the actual trace by assigning probabilities to events that are possibly related to the trajectory of the user [24]. Similar to trace analysis attacks, adversaries trace past movements to determine future locations.

D. OUR CONTRIBUTION

To the best of our knowledge, our previous schemes [25] appear to be the first ones in which blockchain technology has been used in vehicular communication applications. In [25], the security manager network was used to transfer and verify vehicle keys in the across-border requests, rather than forwarding them to the third party authorities. However our previous contributions only focused on VCS security applications, and not at all on the preservation of privacy. We continue our work to use the blockchain structure for privacy preservation. Despite the fact that the paper [4] proposed the concept of shuffling existing pseudonyms by using RSUs first, there is no full explanation on how RSUs run the distribution algorithm. This scheme strongly depends on RSUs, which generates high deployment costs and lacks of robustness in the network [5]. In addition, their system offers digest to record all pseudonym movements, but the digest can be discoursed and misused. Our work is based on the pseudonym shuffling concept. The blockchain technology overcomes the drawback of previous shuffling scheme, which the system frees RSUs and consensus mechanism provides reliable shuffle distribution plan. Due to the nature of blockchain, the digital ledger that contains pseudonym movements stays integrity and authenticity. There are few methods of leader election in blockchains, such as Proof of Elapsed Time [26] and Hedera's Hashgraph [27] – the latter gives us final consistency with probability 1. Our approach is reported for blockchains with Proof of Work but is consistent with using other approaches, although this will require an adjustment of modeling and validation for instances of our schemes.

III. PROPOSED FRAMEWORK

A. SYSTEM MODEL

Nodes in VCS are hierarchically classified into four layers, based on their responsibilities. There are three layers for the service providers, while the service user occupies a single layer [28]. As shown in **Fig.2**, the service provider comprises RSUs, PMs and Public Key Infrastructure (PKI). The PMs and RSUs have wireless communication devices which can communicate over the wireless medium, utilizing VCS communication standards (DSRC [2] or/and C-ITS [29]). RSUs act as access points (APs) which offer interfaces to bridge messages between the service provider and users. Moreover, we assume that each vehicle is required

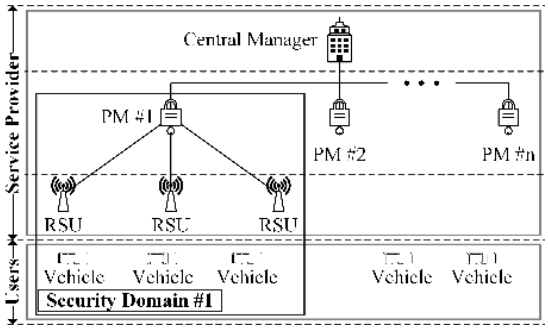

FIGURE 2. VCS network hierarchy.

to be equipped with a built-in computerized device known as an *On-Board Unit* (OBU) – in order to support the VCS standards. A PKI contains a Certificate Authority (CA), an Anonymity Server (AS) and other third-party infrastructure that may support applications.

All the pseudonym-related cryptographic materials, such as anonymous credentials, key pairs and pseudonym certificates are created by the PKI. Each PM has its own logical coverage area, called the *security domain*. PMs help the PKI to manage cryptographic material of security domains that are logically placed below the PKI layer. It is proposed to install PMs in a geographically sparse manner, one for each security domain. Vehicles will transmit and receive safety messages with other vehicles and RSUs. These safety messages are collected by RSUs installed along roads at regular intervals in order to provide maximum network coverage. A safety message includes a pseudonym, a timestamp, and the current vehicle status – including speed, orientation, position, and vehicle dimensions.

Vehicles carry a set of pseudonyms which are used under different time periods in VCS communications. To guarantee privacy, vehicles are supposed to use each pseudonym for only a short duration and frequently switch to a new pseudonym. The US-based VCS standard SAE J2735 [30] defines pseudonym changes to take place within 120 seconds or after 1 km distance travelled (whichever stays longer), while the EU standard ETSI TS 102.867 [31] recommends changing pseudonyms every 5 minutes. The RSUs are equipped with the same network communication technology and are fixed infrastructures with a certain communication coverage area (e.g., a radius of 300 meters in DSRC protocols). The RSUs relay messages between vehicles and PMs, which act as service providers of VCS. To provide context, we compare the traditional and blockchain-based network structures.

1) TRADITIONAL NETWORK STRUCTURE

The traditional structure strictly follows the aforementioned hierarchy. As shown in **Fig.3(a)**, security domains are areas managed by different PMs, and PKIs supervise the network at the top level. A PKI is a trusted authority that provides cryptographic keys, certificates, and long-term identity to

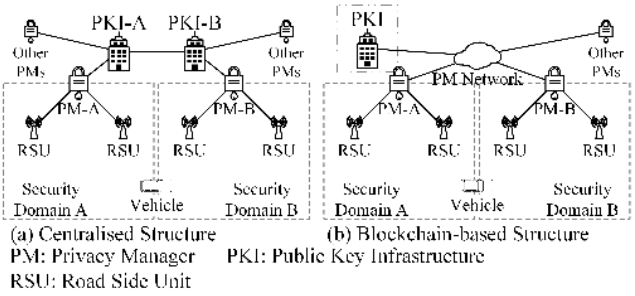

FIGURE 3. Network structures: (a) Traditional (b) Blockchain-based.

all legitimate nodes and infrastructures. Each PKI manages several PMs, as many as are appropriate for the geographical topology of the area. Moreover, PKIs act as bridges that connect different security domains.

Inspired by our previous work [25], we introduce PMs to cover the privacy-related function of VCS. The PM can be seen as the Security Manager (SM) in [25], which has extended privacy protection functions. The RSU is a stationary device placed along roads and at intersections, which is used to gather information about the road traffic and broadcasts it to the OBUs that are within communication range. Also, an RSU can communicate with other RSUs and the CA to exchange messages related to the road traffic through a secure channel. Our previous work [32] follows the traditional network structure as do most other works, such as [33], [34].

2) BLOCKCHAIN BASED STRUCTURE

The PMs manage a certain amount of RSUs based on the geographic distribution of RSUs, shown in **Fig.3(b)**. In contrast to a traditional network structure, a PKI is isolated and would be a part of an existing authority such as a Driver and Vehicle Licensing Agency. The PKI is designed to generate specific cryptographic credentials for all the nodes and to link vehicles to their long-term identities. Cryptographic credentials – such as vehicle identities, pseudonyms and pseudonym certificates – are supposed to be kept in a secured facility to fulfill privacy and security requirements [35]. Thus the central managers are accessed in the following two situations:

(i) *Initial Registration:* New vehicles need to apply for initial registration when they leave the manufacturer and participate in a new security domain for the first time. **(ii)** *Adversary revocation*. In the blockchain-based structure, malicious behaviors are recognized through using blockchain look-up. The identity (including pseudonyms) of the adversary is then publicized, once the malicious behaviors have been confirmed.

As a result, our proposed blockchain-based structure could enable PMs to securely keep all communication logs without reliance on a central party. All PMs are connected with each other and the PKI on a domain. PMs communication mainly contains peer-to-peer pseudonym sets exchange, encapsulated in transactions. Similar to Bitcoin, the ledger keeps all transactions from the beginning. And PMs act as miners to put transactions into a block within a fixed period of time. With this blockchain-based structure, our system can reuse pseudonyms by shuffling them between PMs. The shuffle results will be determined by the first miner and be added to the block. Hence a blockchain can be maintained for the purposes of pseudonym management. We also made assumptions for the blockchain structure:

Assumption 1 (Role of Miners): Generally speaking, nodes are classified into two roles according to different responsibilities among the blockchain network, namely service user and miners. The miners are nodes with powerful computation power who use their computation power to maintain the blockchain. In the Bitcoin network, nodes decide on their own whether or not they want to take on the role of a miner. Bitcoin pays the miner who wins the mining race for the next block a reward, in addition to transaction fees embedded in that block. This creates incentives that ensure that mining takes place, but also causes problems such as dramatic increases in difficulty when Bitcoins become very valuable in fiat currencies.

In our blockchain-based scheme, we assume all the block mining tasks are carried out by all the PMs as procured resources, and so they do not need any incentives and won't necessarily receive rewards – as discussed in [36] previously. This is sensible in our setting because we believe that pseudonym management, as part of ITS management, should be run by the appropriate organization of the government (e.g. the Driver and Vehicle Licensing Agency in the UK). All the PMs take the roles of service user and miner at the same time. It may also be attractive to use Proof of Kernel Mode [16] as a variant of Proof of Work that randomly and securely would select an expected number of PMs for mining each time. This will then allow for using a lower level of difficulty and will save costs as only those nodes selection for the next mining race will consume energy in mining.

Assumption 2 (Approximate Mining Synchrony): It is beneficial to be able to ensure that all the PMs start mining tasks at approximately the same time. As the navigation service is contained in the ITS applications, each vehicle should have a synchronized clock. This helps to limit the deadline for each transaction collection interval. Any lack of synchrony may also be contained by using a combination of, for example, Proof of (Kernel) Work and Proof of Elapsed Time, as discussed and modeled in [16].

Assumption 3 (Consensus): Proof of Work is the only consensus mechanism that has been tested successfully and in a sustained manner in a highly adversarial environment, and is the only known cryptographic puzzle that meets these testing requirements. Alternative consensus mechanisms such as the ones aforementioned have not yet been tested in real and adversarial practice. This is why we favor PoW-style consensus given that an ITS is part of a regional or national critical infrastructure that may be subject to aggressive attacks, perhaps even facilitated by compromised insiders. PoW gives us this resiliency even against corrupted PMs and low levels of difficultly, especially when used with a Proof of Kernel

version of PoW, which give a balance of security and shorter compute time at lower cost.

B. THREAT MODEL

Due to fact that broadcast safety messages need to be sent, an eavesdropper may track a specific vehicle and monitor its location information by leveraging these periodic safety messages [10]. In this paper, we consider external and internal attacks. The two types of external attacks are *global passive attack* and *local passive attack*. The two types of internal attacks are *internal tricking attack* and *internal betrayal attack*.

1) GLOBAL PASSIVE ADVERSARY (GPA)

A global adversary has the overall coverage of a connected vehicle network. This GPA can locate and track any vehicle in any region of interest by eavesdropping its broadcast beacon messages.

2) LOCAL PASSIVE ADVERSARY (LPA)

The local passive adversary is limited in its location tracking capability in a region of interest, since it can only exploit the deployed infrastructures for eavesdropping and estimating locations of vehicle broadcasts. Hence, the region over which the LPA can track vehicles is dependent on the vehicle transmission range.

3) INTERNAL BETRAYAL ADVERSARY (IBA)

An internal adversary is a compromised node that becomes an adversary in the network system. The internal attacker could spoof safety massages and collude with a global passive attacker or local passive attacker to track a target vehicle. After swapping or obtaining privacy-related information (e.g., the pseudonyms) with the target vehicle, the malicious user can leak the information to the global passive attacker or local passive attacker to link the target's location and real identity.

4) INTERNAL TRICKING ADVERSARY (ITA)

Unlike the IBA, the internal tricking adversary will use pseudonyms which have been allocated to others, allowing it to confuse the vehicular network system and to attack other nodes.

There are other methods for attacking the vehicular network system. For example, accessing traffic monitoring cameras or hijacking the Global Positioning System (GPS) allows tracking the target vehicle. Furthermore, adversaries may be able to compromise privacy managers to attach a false block into the blockchain. Yet acquiring either of these capabilities, access to a traffic monitor that controls national traffic operations centre or taking control the blockchain itself, requires a significant effort – e.g. having at least 51% of the total blockchain network's processing power.

C. PSEUDONYM MANAGEMENT

We now introduce our blockchain-based pseudonym management scheme, which intends to reuse pseudonyms and address the distribution issue that decentralized systems have regarding pseudonym shuffling. The main symbols used in this scheme are listed in **Table.1**.

TABLE 1. Symbols used in the paper.

Notation	Description
PM_x	The x^{th} PM.
PK_x, SK_x	The public and private keys for the vehicle x.
$PN_i^{PM_x}$	The i^{th} pseudonym from PM_x.
PM_x	The x^{th} PM.
n_T	the number of transactions.
$\mathbb{R}$	a set of all possible number of transactions.
$cipherinfo$	Encrypted infomation of pseudonym sets with public key of destination PM.
$Sig\{CI\}_{SK_{PM-s}}$	Digital signature on CipherInfo with private key of the source PM.

1) PSEUDONYM DISTRIBUTION

From a management perspective, pseudonym sets for each car that are presently stored in the OBU will be depleted. Authors in [4] mentioned that the use of a backbone network of RSUs may resolve the aforementioned issues and reuse pseudonyms for a limited period of time and in different geographic areas. However, the distributed nature of these systems then creates an additional optimization problem: it is hard to balance the volume of incoming and outgoing pseudonyms without a centralized means of controlling this. This issue remains even when using distributed versions of the simplex algorithm in order to alleviate the computational demands on the optimization problem.

In terms of Privacy-by-Design for the VCS network, we should consider pseudonym generation and distribution more wisely. Specifically, the number of pseudonyms generated by a PKI should be limited but sufficiently large in order to meet demand. Two initialization events are introduced to finish the entire initialization stage, namely the permanent identity and pseudonym generation. The permanent identity contains the identity number *ID*, a certificate *CERT* and key pairs (private key *SK* and public key *PK*) which are used to prove the real node identity or the initial registration identity. These credentials are generated by PKIs and distributed to the manufacturers who are responsible for producing vehicles and the VCS infrastructure.

The distribution procedure between a PKI and manufacturers is finalized via highly secured connections, such as optical fiber or cable connections. PKIs generate a certain number of pseudonyms offline and then distribute pseudonym sets to each PM. Each pseudonym set $\{id_1 \cdots id_n\}$ contains the corresponding pseudonym certificates $\{cert_1 \cdots cert_n\}$ and encryption private/public key pairs $\{sk_1/pk_1 \cdots sk_n/pk_n\}$. The number of pseudonyms inside sets is determined by the density of traffic in corresponding areas which all RSUs reported to their PMs. Pseudonyms will be encrypted and

signed to maintain secrecy before a PKI distributes them to PMs. The encryption and signing use the public key PK_{PM} of the PM and secret key SK_{PKI} of the PKI, respectively. To summarize:

i. *Generates Permanent Identity:*

PKI *generates ID, CERT, SK&PK*

ii. *Distributes Permanent Identity:*

PKI *sends ID, CERT, SK&PK to* **Manufacturers** :

$\{ID + CERT + SK + PK\}_{secured\ channel}$

Manufacturers *issues ID, CERT, SK&PK to* **Vehicles** *or* **Infrastructures** :

$\{ID + CERT + SK + PK\}_{file\ transfer}$

iii. *Generates Pseudonyms:*

$PKI\ generates: \{id_1 \cdots id_n\}, \{cert_1 \cdots cert_n\}\ and\ \{sk_1/pk_1 \cdots sk_n/pk_n\}$

iv. *Distributes Pseudonyms:*

PKI *sends pseudonyms to* **PM** :

$\{id_1 \cdots id_n\}_{PK_{PM}} + \{cert_1 \cdots cert_n\}_{PK_{PM}} + \{sk_1/pk_1 \cdots sk_n/pk_n\}_{PK_{PM}} + Signature_{SK_{PKI}}$

2) PSEUDONYM SHUFFLING

To keep sufficient pseudonyms to allow frequent changing across vehicles, PMs are responsible for retrieving used pseudonyms and issuing fresh pseudonyms. There are two challenges for this shuffling scheme: (1) the path of pseudonym exchanges needs to be protected; otherwise, the attacker could subject the path to further analysis in order to constrain the possible pseudonyms delivered to vehicles in certain RSUs' ranges. (2) The demand of pseudonym for each privacy manager is supposed to be fulfilled. For instance, the PM covers central London would need more pseudonyms than PMs in countryside because different locations have different traffic. We use blockchain technology to deal with these challenges for pseudonym management, as it could provide sufficient randomness on the shuffling path and enough computation power to tackle the distribution optimization problem.

When vehicles operate on a road, they will frequently change pseudonyms based on a certain pseudonym change algorithm. For our pseudonym management, pseudonym changes are supposed to execute within mixed zones, which are geographic regions within the VCS environment as shown in **Fig.4.** Generally speaking, the mixed zone must be selected carefully in order to maximize the level of privacy. For example, traffic junctions, roundabouts and temporary car parks will help a lot with mixing privacy-related messages as they contain a large number of vehicles with similar status.

Algorithm.1 briefly describes the mechanism used when a vehicle joins a mixed zone. We propose that traffic junctions and roundabouts could be treated as physical mixed zone

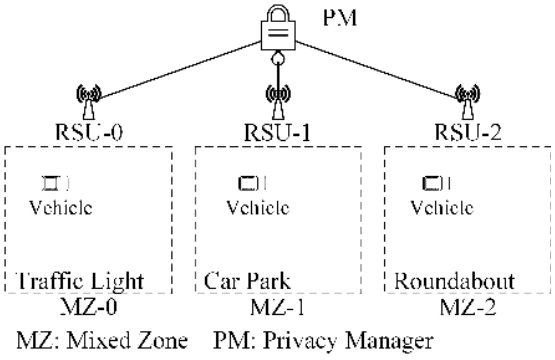

FIGURE 4. Mixed zone example.

Algorithm 1 The Joining-Mixed-Zone Mechanism

Input: : Current PM id PM_x, Public Key of PM_x: PK_x, used pseudonym set PN_{used}, a **Mixed Zone** area under managed by PM_x, Location Cloaking Requirement of **Mixed Zone**: $Cloak\{\}$, Current location $Location$

if (Vehicle enters a RSU cover area) **then**
    Mixed Zone = **True**
else if (Vehicle enters a virtual mixed zone) **then**
    Mixed Zone = **True**
else
    Mixed Zone = **False**
end if
if (**Mixed Zone** = **True**) **then**
    Cloaks the location information $Cloak\{Location\}$;
    Broadcasts safety messages using cloaked location;
    Encrypts pseudonym by PM's public key: $Enc\{PN_{used}\}_{PK_x}$;
    Sends $Enc\{PN_{used}\}_{PK_x}$ to PM_x;
end if
End Algorithm

where RSUs can be placed, while traffic lights or other places at which enough vehicles could gather in close proximity are seen as ''virtual mixed zones''. In virtual mixed zones, vehicles would trigger pseudonym change even when the vehicles are on a highway with vehicles of a similar status (e.g., similar speed, same heading, and so forth). A vehicle first cloaks its location information according to the specific cloaking algorithm of a mixed zone. This aims to mix all vehicles so that the probability of tracking can be minimised.

Initially, vehicles carry a set of pseudonyms installed at the time of vehicle manufacture. A vehicle marks a pseudonym as ''*used*'' and switches to a new one if the pseudonym meets its expiry conditions. We defined a threshold for used pseudonym sets, a fixed percent of the number of pseudonyms.

To assure the vehicle will not only have enough new pseudonyms to use after it gave up used pseudonyms but that it also collects a maximum number of pseudonyms in order to reduce transmission overhead, the threshold is set to cover

TABLE 2. The format of forwarded package.

Package Header			Payload
Type	PM no.	PN number	Pseudonyms
Privacy	PM_1	n_1	$\{PN_1^{PM_1}, PN_2^{PM_1} \cdots PN_{n_1}^{PM_1}\}$
Privacy	PM_2	n_2	$\{PN_1^{PM_2}, PN_2^{PM_2} \cdots PN_{n_2}^{PM_2}\}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$
Privacy	PM_i	n_i	$\{PN_1^{PM_i}, PN_2^{PM_i} \cdots PN_{n_i}^{PM_i}\}$

the majority of pseudonyms. The vehicle then encapsulates all used pseudonyms into a package that it sends to the current RSU. PMs will collect used pseudonym packages for a fixed period of time from all RSUs that are situated in its range and then aggregate all packages into a single transaction. All packages and transactions are signed by their senders and encrypted with the receiving PMs' public keys. Hence PMs could assure all pseudonyms are integrated and authenticated. Then the PMs upload all used pseudonyms, related indexes, and the number of pseudonyms in the *PM cloud*. After that, each PM in the network will make a copy of all pseudonyms that have been uploaded for this shuffle. Since every communication between PMs contains timestamps, pseudonym shuffle will be triggered in every fixed interval. When pseudonym shuffling commences, all PMs pull the demand of each PM from the cloud and add those demands to their own list. The PMs will randomly choose pseudonym sets and allocate them to every PM based on the number of required pseudonym sets.

Algorithm 2 The Pseudonym Shuffling Scheme

1: **for** $(x = 1; x \leqslant i; x{+}{+})$ **do**
2: PM_x gathers all the used pseudonyms from mixed zones it manages;
3: $PN^{PM_x} = \{PN_1^{PM_x} \cdots PN_{n_x}^{PM_x}\}$;
4: Counts the number of used pseudonyms $= n_x$;
5: Encapsulates PN^{PM_x} into package and sends into PM cloud network;
6: **end for**
7: **for** $(x = 1; x \leqslant i; x{+}{+})$ **do**
8: PM_x picks up all the pseudonym package within PM cloud network;
9: **Shuffles** the pseudonym sequence and **relocates** to destination PMs;
10: **end for**
11: All the PMs start **Mining**;
12: The mining winner broadcasts the **Block** into PM network;
13: **for** $(x = 1; x \leqslant i; x{+}{+})$ **do**
14: Retrieves new pseudonyms for PM_x;
15: **end for**
16: **End Algorithm**

The shuffling algorithm is outlined in **Algorithm.2**. **Table.2** illustrates an example of all the forwarded packages within the PM cloud network of i many PMs, ranging from PM_1 to PM_i. The first field in the package header indicates the type of this packet, used for further extending the service to security applications. The remaining fields in the header are the PM number $\{n_1 \cdots n_i\}$ and the number of pseudonyms which are donated from the PM, respectively. The payload field contains all the used pseudonyms PN^{PM}.

An example of the shuffle mapping result is shown in **Table.2**. Here we assume random variables $\{a, b, c, d\} \in [1, i]$ and $aa \in [1, n_1], bb \in [1, n_2], cc \in [1, n_3], dd \in [1, n_4]$. The first line means a PM selects a previously-used pseudonym $PN_{aa}^{PM_a}$ from PM_a. This source to destination result is marked by the sequence number 1. After creating this list, each PM encapsulates its list into a Blockchain-based transaction. Then PMs will try to mine for consensus, e.g. by calculating Proof of Work (PoW). Whoever first finishes that mining race must add the mined block into the Blockchain. All PMs will validate such a new block and, if validated, follow the block's description of how to allocate pseudonym sets. Since each transaction is signed by the sending PM's private key and encrypted by the receiving PM's public key, each transaction is only visible to relative PMs. Even though all transactions are attached in the block and the block was broadcast to all PMs, others who neither sent nor received a specific transaction cannot obtain any information from that transaction. Hence each PM can only decrypt its own transactions (for which they were the receiver). So each PM will perform the pseudonym shuffle by shuffling all pseudonym indexes individually. After each shuffling, all PMs will delete all copies of pseudonym sets. The format of transactions and the mining will be described next.

TABLE 3. Shuffle mapping table.

Src	Dest	Seq no.	Pseudonyms
PM_a	PM_1	1	$PN_{aa}^{PM_a}$
$\cdots$	PM_1	$\cdots$	$\cdots$
PM_b	PM_1	n_1	$PN_{bb}^{PM_b}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$
PM_c	PM_i	$\sum_{k=1}^{i} n_k - n_i + 1$	$PN_{cc}^{PM_c}$
$\cdots$	PM_i	$\cdots$	$\cdots$
PM_d	PM_i	$\sum_{k=1}^{i} n_k$	$PN_{dd}^{PM_d}$

3) TRANSACTION FORMAT

The transaction ledger is designed to encapsulate pseudonym materials from a source-privacy manager to a destination-privacy manager. An example of the transaction ledger is shown in **Table.3**. In the ledger, the left-hand side shows the source PM address and the destination PM address, while the right-hand side includes the pseudonym sets of this transaction and their corresponding credentials (e.g., key pairs and certificate). The data in each column of the payload on the right-hand side of the table has been encrypted by the public key of the destination PM $PK_{PM-dest}$, which establishes the transaction's integrity and confidentiality. Only the destination PM who has the private key $SK_{PM-dest}$ can decrypt this

information. Messages are encrypted with the private key of the source PM $SK_{PM-sour}$. With the use of a digital signature, the encrypted transaction information can be prevented from spoofing attacks and eavesdropping by malicious users – since they would need to forge signatures.

TABLE 4. The format of transaction.

Transaction Header
Hashed result of the transaction
Number of this transaction in block
Current privacy manager address *PM-sour*
Destination privacy manager address *PM-dest*
Signature of this transaction to ensure integrity and authentication $Sig\{CipherInfo + PMdest\}_{SK_{PM\text{-}sour}}$
Payload: (Encrypted Transaction Information) $Cipherinfo = En\{Pseudonym\ sets\}_{PK_{PM-dest}}$

Table.4 shows the format of each transaction in the ledger, containing transaction header and payload. In the transaction header, the number of this transaction specifies the position at which this transaction is located in the ledger. The source and destination PM address are similar to Bitcoin inputs and outputs seen in [6]. The signature occupies the last position of the transaction to maintain the authentication, integrity, and non-repudiation of key-transfer information. The *cipherinfo* has already been discussed above.

TABLE 5. The format of block.

Block Header	
Field	Description
Version	Block Version Number
Previous Block Hash	Hash of the previous block in the chain
Merkle Tree Root	Hash of the merkle tree root $Root_M$
Timestamp	Creation time of this block
Targeted Difficulty	The Proof-Of-Work difficulty target
Nonce	A counter for the Proof-Of-Work
Block Payload (Transactions)	
Transaction No.1 $\cdots$ Transaction No.n	

4) BLOCK FORMAT

A block is designed to store all transactions as ledgers. All blocks need to be joined to form a large chain — the *blockchain*. The format of a block is shown in **Table.5**. The first row shows the block number, which is the sequence number of the block within the entire chain. The hash of the previous block securely links this block to its parent one through the mining process. This makes is extremely hard to replace contiguous sub-chains with other data and to convince other nodes of the validity of such changes. The Merkle tree root is used for securing the integrity of transactions within a block [37]: all transactions in this block are jointly authenticated into the Merkle tree root, so that any alteration on any transactions would cause a different value of Merkle root value. As in Bitcoin, we add a timestamp to prove when this block of transactions was created and to prevent time tampering. The fields for targeted difficulty and nonce are designed for Proof of Work, which creates a digital receipt of which first node mined that block. The mining process and Proof of Work for our approach is described in the next section. The payload field contains the aforementioned transactions that the block creator randomly allocated.

5) CONSENSUS ALGORITHM

The consensus mechanism used in a blockchain establishes, in a distributed way, an agreement between all network nodes, instead of relying on a central party's decision. The most widely known and used consensus mechanism for blockchains is Proof of Work (PoW), which is a mining race in which nodes try to solve a hard cryptographic puzzle concurrently. The PoW system was originally proposed as a means of deterring spam email [38]. All PoW-based applications (e.g, Bitcoin and the current Ethereum) require participating nodes to contribute a significant amount of computation power in order to obtain a digital proof of work that can be verified easily. The process by which nodes compete in finding such proof of work is called *mining*. The first node to solve the cryptographic puzzle for the next block to be added to the chain will be elected/accepted as leader and is then able to add the new block for which it found proof of work to the chain.

In blockchain-based applications, a cryptographically strong hash function is used to calculate the proof of work. In our case, we use double SHA-256 on the previous block hash result and the Merkle tree with related time stamp of the new block as input to that hash function.

The proof of work involves adding some random information, a nonce value, to that input until the resulting hash has a desired minimal number of leading 0 bits. Consequently, proof of work has several desirable features. For example, a miner that has had k failed mining attempts has no advantage in the $k+1$th attempt in comparison to another miner who just begins its first attempt of solving PoW. Also, PMs are very likely to have different mining times due to the exponential distribution for the expected time to find proof of work within a certain period of time. Also, we assume that all PMs randomly generate transactions that result in different root hashes and that they all use the same hardware specification, making this a blockchain system in which miners are procured resources as proposed in [36]. Therefore, all PMs have the same probability of getting the correct hash results within a certain period of time t, assuring the randomness of the resulting shuffling.

Since all PMs contain identical processing modules and since they are assumed to link with highly secured wire connections, we may set the level of difficulty (the number of leading 0 bits in the hash output) required to be rather low. This low level of difficulty allows for a short Proof of Work computation time, resulting in an efficient yet resilient consensus mechanism.

6) SHUFFLE TIME COMPOSITION

Table.6 demonstrates each time factor for the shuffling process. The variable t_{prep} is the time needed for preparing a block, including the PM's generation of a randomized transaction ledger and the time cost of block preparation.

TABLE 6. The time elements of processing procedures.

Parent Field	Description of Parent Field	Child Field	Description of Child Field
t_{prep}	The time cost to prepare block which will be mined later	t_{rand}	Calculation time to generate random transactions
		t_{fill}	Time cost to insert transactions into the block message
		t_{merkle}	Calculation time to get Merkle Tree Root
		t_{header}	Processing time to prepare block header
$t_{transfer}$	Transmission time cost in PM network	t_P	Propagation time in network cable
$t_{processing}$	Processing time for message **Verification**	t_V	Processing time to verify signature
t_{mining}	Mining time for shuffle process	t_M	The average time to mining a block

We denote by N the number of PMs. To calculate the total time cost of the shuffling process, we also need the number of transactions ($n_T \in \mathbb{R}$) where $\mathbb{R}$ is a set of all possible number of transactions. Given that,

$$\mathbb{R} = 2 \times (n-1) \times n \tag{1}$$

where $n \in \mathbb{Z}^*$ and $\mathbb{Z}^* = \{0\} \cup \mathbb{Z}^+$. $\mathbb{Z}^+$ denotes the positive integers. Hence the total time cost can be described as:

$$t_B = n_T \times t_V + 2 \times t_P + t_{prep} + t_M \tag{2}$$

So the total time cost can be expressed as seen in equation (2), that contains all time factors. Note that the total transaction verification time ($n_T \times t_V$) depends on the number of transactions (n_T). The preparation time and mining time are added to reflect the time needed for creating a block in the blockchain.

IV. SYSTEM EVALUATION

A. PRIVACY ATTACK AND DEFENCE ANALYSIS

Researchers such as Yu in [39] state that the power of identity and location privacy preservation in pseudonym-based systems is determined by the unpredictability of mapping temporary identifiers (pseudonyms) to vehicular permanent identities. Accordingly, our blockchain based pseudonym-management system aims to improve the unpredictability of pseudonym mixtures while at the same time reducing the cost and effort of constantly generating new pseudonym certificates by shuffling used pseudonyms. Privacy attacks pose a serious issue in current ITS that require addressing. Without proper identity and location-privacy preservation, attacks, such as vehicle tracking, location manipulating and so forth could cause serious damage to vehicles and compromise the safety of human actors. Moreover, the lack of such abilities will hinder the development and acceptability of the Internet of Connected Vehicles.

In the following, we show how our approach can address some pertinent attacks and the defence measures.

1) GPA AND LPA

The most common privacy attack is when an adversary passively eavesdrops vehicles' beacon messages. Other than the difference of coverage, both GPA and LPA could obtain the timestamps and location of the joining and leaving of vehicles in order to derive a likelihood distribution over possible mappings. As mentioned in Section II, several works claim that they can predict vehicles' trajectories with a brute-force collection of beacon messages even when vehicles change pseudonyms frequently. In contrast, our proposed system not only allows vehicles to change pseudonyms simultaneously at a mixed zone, but also at the virtual mixed zone as long as there are sufficiently many vehicles with similar context within that zone. In this case, for both GPA and LPA, the unpredictability of mapping vehicles is accumulated.

2) IBA AND ITA

As already stated above, we focus on two specific internal adversaries, namely internal betrayal adversary and internal tricking adversary. Whenever the internal betrayal adversary (IBA) obtains a pseudonym of the target vehicle, it is able to perform privacy attacks on vehicles (e.g., to manipulate safety messages with the temporary identity of the target vehicle) or share the information to the global passive adversary so that the pseudonym of the target vehicle could be mapped to its real identity. In contrast, our proposed scheme prevents the IBA from accessing others' pseudonyms simply, as vehicles will not exchange pseudonyms with each other. According to the pseudonym change scheme of the proposed system, vehicles only update pseudonyms from their own pseudonym sets which have been allocated by the RSUs (not vehicles). After acquiring a new pseudonym set from the RSU, the vehicle cannot retrieve the original source of the new pseudonyms in that set. Therefore, the IBA could not obtain any useful information from its surrounding or related vehicles.

In terms of ITA, the malicious user will keep and repetitively use pseudonyms that have been uploaded to RSUs and allocated to other vehicles. While other vehicles are using the same ones, the ITA could use the pseudonyms to confuse the vehicular network system and launch other attacks. To deal with this problem, the system behaves as follows. If the adversary stays in the current RSU's coverage, the RSU will realize that the adversary keeps using the old pseudonyms. Then the RSU will mark the vehicle as adversarial and broadcast this information to other vehicles. If the adversary leaves the RSU's coverage, no one other than a CA could know that the attacker is a ITA, due to the feature of the blockchain based shuffling system: each PM will only recognize its own related pseudonym shuffle routes from transactions of the block and will not be able to see (unencrypted) other transactions in the block. So RSUs and PMs will not

know the allocation of each pseudonym and also could not decide whether the adversary is using false pseudonyms. However, once the attacker launches other attacks in the false pseudonyms, such as spoofing messages, that can be detected, the CA will retrieve transactions of the blockchain. Hence it can spot the adversary and perform revocation of original credentials.

3) COMPROMISED PM

The privacy manager is a crucial part of this pseudonym-management system. Therefore, we normally assume that PMs are relatively secure, most likely run and maintained by government agencies or similar governing bodies. In addition, blockchains are well known for providing high robustness and for being hard to manipulate by an adversary. However, let us consider a worse circumstance in which one PM is compromised by a malicious user. There are several attack scenarios that may be enabled by this. However when a PM is compromised or has lost connection with the blockchain network, the whole blockchain will discard the PM after repeated failed attempts to acquire its response. In addition, all pseudonym sets that this PM received from previous round pseudonym shuffles will be abandoned from the PKIs and will thus not be used again.

4) SPOOFING BLOCK ATTACK

The spoof block attack assumes that a privacy manager (PM) has been compromised or betrayed so that it is broadcasting false blocks into the blockchain cloud. Then an adversary can re-arrange pseudonym allocations and manipulate vehicles' identity. But in order to consistently send out forged blocks and to have them accepted by the blockchain network, the compromised PM will need to have at least 51 % computation power of the total blockchain based network since Proof of Work is used as consensus mechanism. Otherwise, this PM will not be able to control the mining process and so won't be able to determine the history and future of the blockchain.

B. PERFORMANCE ANALYSIS

We now offer a quantitative analysis of our approach.

1) SIMULATION ASSUMPTIONS

The simulation of the pseudonym management scheme was carried out using OMNET++ with the dedicated simulation package (Veins and PREXT) [40]. Elliptic Curve Integrated Encryption Scheme (ECIES) [41] with elliptic curve secp160r1 in Crypto++ [42] is selected not only for cryptographic scheme ECIES, but also for the Elliptic Curve Digital Signature Algorithm (ECDSA) as well. We simulated our proposed scheme on our desktop machine equipped with an Intel Core i7, 8 GB RAM, and a display card Inter HD Graphics 530. Our simulation considered 300 vehicles. All vehicles followed the pseudonym change scheme of our proposed system. We set the traffic density at 50 vehicles per kilometre and the transmission range of the target at 50 meters based on [4], [32]. The performance results are broken into four parts. Firstly, we evaluate the pseudonym reuse frequency. Secondly, we calculate the total amount pseudonym usage and compare with EU ETSI standard. Then the degree of anonymity is studied compared to other existing schemes. Lastly, we investigate the time cost of the entire process.

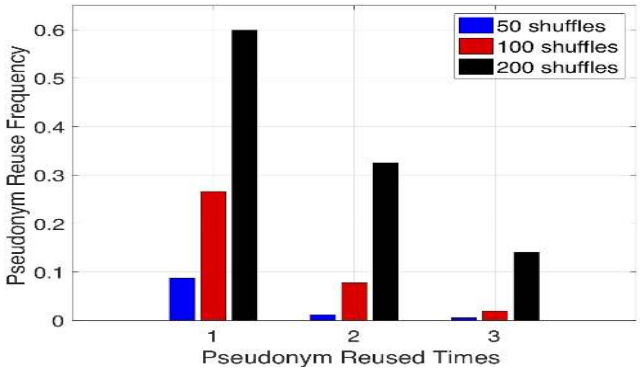

FIGURE 5. The percentage of pseudonyms re-usage vs. pseudonym reused times.

2) PSEUDONYM REUSE FREQUENCY

We first study the pseudonym reuse frequency to demonstrate shuffling effectiveness. We let each vehicle carry 10 pseudonyms with a threshold setting of 8, meaning whenever a vehicle has 2 unused pseudonyms it will upload 8 used pseudonyms to its RSU. We simulate the scenarios in which the shuffling process happens 50, 100 and 200 times. Fig.5 shows the percentage of how many pseudonyms have been reused over 1, 2 and 3 times respectively. As can be seen from the graph, 60% of the pseudonyms have been reused at least once when the shuffling process performs 200 times. In addition, there is a significant drop in the frequency of reused pseudonyms when the number of used times increases. Hence the results of pseudonym reuse frequency indicate that our scheme assures pseudonyms could be reused in different locations in limited shuffling iterations. Despite the fact that the increased reuse times of each pseudonym could free up more storage for OBU, the results also show that the number of pseudonym reuse times affects the percentage of reused pseudonyms. Only 0.15% of the pseudonyms have been reused over 3 times in 50 shuffling iterations. Therefore, we need to have a reasonable understanding of the suitable number of shuffling times used to measure the anonymity performance of our scheme.

3) PSEUDONYM TOTAL AMOUNT

To quantify the efficiency of reusing pseudonyms, we compare the total amount of pseudonyms that the proposed system needs for a 24-hour period with the ETSI standard [43] for the change frequency. Since the ETSI standard suggests vehicles change pseudonyms every 5 minutes, one vehicle needs 288 pseudonyms for 24 hours. For the proposed scheme, we denote that the capacity of storing pseudonyms

of each vehicle is X. Based on the size of storage in OBU, vehicles currently could have from 10 pseudonyms to nearly 1,000 pseudonyms [44]. We continue use 100 vehicles in this comparison. The system following the ETSI standard would need over $288 \times 100 = 28,800$ pseudonym certificates each day and 864, 000 each month, whereas the proposed scheme uses 100, 000 pseudonyms even with a storage capacity of 1, 000 pseudonym certificates until the system decides to replace new pseudonyms. In fact, our scheme is not affected by time duration and pseudonym certificates would be re-used over time.

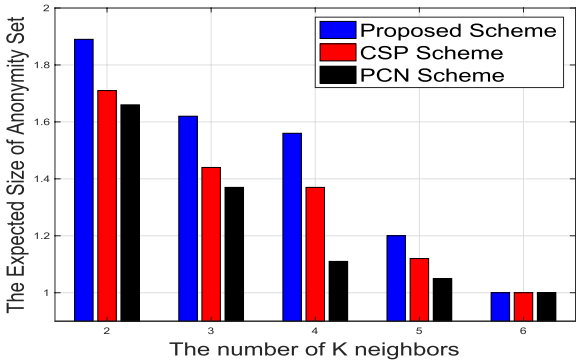

FIGURE 6. Expected anonymity set size with different value of *k* neighbors.

4) ANONYMITY SET SIZE

In this simulation, we run the shuffle process 200 times, based on the results of pseudonym reuse frequency above, in order to achieve higher performance. Fig. 6 indicates the influence of the k neighbors on the expected size of the anonymity set. The size of anonymity set is a measure of the level of anonymity provided by the cloaking algorithm, normalized by the level of anonymity required by the messages. Note that the relative anonymity level cannot go below 1. Higher anonymity set sizes mean that messages are anonymous with larger k values than the user-specified minimum k-anonymity levels. we calculate the maximum AS size encountered by each vehicle and then taking the average over all vehicles. The maximum AS size of a subject vehicle is obtained by finding the maximum number of nearby vehicles, including itself, that changed their pseudonyms simultaneously with a pseudonym change by this subject vehicle. Two vehicles are considered nearby if they are located within a distance of 100 m. As can be seen, we compare the proposed scheme with the coordinated silent period (CSP) scheme and with the cooperative pseudonym change (PCN) scheme that we mentioned in our discussion of related work [8], [9]. CSP proposes a approach that all vehicles in certain area completely cease any communication and changes its pseudonym for a period of time to maximize the anonymity. PCN illustrates that the target wait till k neighbors around to change pseudonyms together. Our proposed scheme achieves a better level of anonymity as the expected anonymity set size is greater. In addition, the expected anonymity set sizes have significant drops in our proposed scheme and other schemes when the value of k increases. This is the case since it is less probable to find greater or identical k neighboring vehicles when the value of k becomes large. When k is set to 6, the anonymity set sizes of all three schemes are equal to 1, which means that vehicles can only find less than k neighbor's.

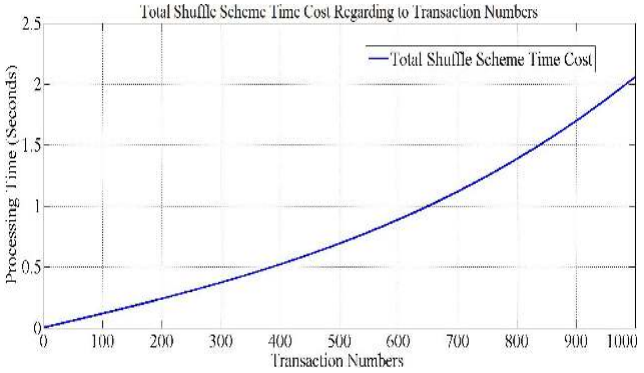

FIGURE 7. The total time cost regarding to transaction numbers.

5) PROCESSING TIME

We illustrate the total processing time of our proposed scheme. We acquire each time component from **Table.6** and calculate the result of total time cost based on equation (2). As can be seen in Fig.7, the total time increases when the transaction number grows. Due to the benefits that come with our design of privacy managers, the transaction number is limited as equation (1) demonstrated. Therefore, the total shuffle process time stays within a reasonably short period. For instance, we take a medium size city as an example. We assume 30 privacy managers are placed in the city, and each of them covers several RSUs. Based on equation (1), the number of transactions can be less than 100 in off-peak hours, while the maximum number could exceed 1000 in peak hours. The total processing time varies from 0.2 seconds for 100 transactions to over 2 seconds for 1,000 transactions.

V. CONCLUSION

In this paper, we have proposed a novel decentralized pseudonym-management scheme for the Internet of Connected Vehicles that makes use of a blockchain based on Proof of Work in order to make the overall system more resilient to known privacy and security attacks. The proposed scheme provides a method to effectively manage pseudonyms from distribution and re-utilization, and a pseudonym change-scheme that combines physical mixed zones with virtual mixed zones. The paper discussed several types of vehicle privacy attacks and defence measures that are enabled by our proposed blockchain-based system. We used OMNET++ and Veins to quantitatively evaluate the proposed scheme. The simulated pseudonym reuse frequency and total amount of pseudonym consumption corroborate that our proposed scheme can be used in Connected Vehicular Networks and that it could significantly reduce

the cost of pseudonym related generation and maintenance of credentials. Moreover, the simulation results show that the scheme achieves better anonymity than existing schemes when the shuffling process on pseudonyms is performed 200 times. In addition, a total process time is computed which shows that our scheme is capable of performing pseudonym shuffling with over 1, 000 blockchain transactions in 2 seconds.

REFERENCES

[1] Y. Fangchun, W. Shangguang, L. Jinglin, L. Zhihan, and S. Qibo, "An overview of Internet of vehicles," *China Commun.*, vol. 11, no. 10, pp. 1–15, Oct. 2014.
[2] J. B. Kenney, "Dedicated short-range communications (DSRC) standards in the United States," *Proc. IEEE*, vol. 99, no. 7, pp. 1162–1182, Jul. 2011.
[3] B. Brecht, D. Therriault, A. Weimerskirch, W. Whyte, V. Kumar, T. Hehn, and R. Goudy, "A security credential management system for V2X communications," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 12, pp. 3850–3871, Dec. 2018.
[4] H. Artail and N. Abbani, "A pseudonym management system to achieve anonymity in vehicular ad hoc networks," *IEEE Trans. Dependable Secure Comput.*, vol. 13, no. 1, pp. 106–119, Jan./Feb. 2016.
[5] A. Boualouache, S.-M. Senouci, and S. Moussaoui, "Towards an efficient pseudonym management and changing scheme for vehicular ad-hoc networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–7.
[6] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Tech. Rep., 2008. [Online]. Available: http://www.bitcoin.org
[7] F. Dötzer, "Privacy issues in vehicular ad hoc networks," in *Privacy Enhancing Technologies*, G. Danezis and D. Martin, Eds. Berlin, Germany: Springer, 2006, pp. 197–209.
[8] A. Tomandl, F. Scheuer, and H. Federrath, "Simulation-based evaluation of techniques for privacy protection in VANETs," in *Proc. IEEE 8th Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, Oct. 2012, pp. 165–172.
[9] Y. Pan and J. Li, "An analysis of anonymity for cooperative pseudonym change scheme in one-dimensional VANETs," in *Proc. IEEE 16th Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, May 2012, pp. 251–257.
[10] K. Emara, W. Woerndl, and J. Schlichter, "Vehicle tracking using vehicular network beacons," in *Proc. IEEE 14th Int. Symp. Workshops World Wireless, Mobile Multimedia Netw. (WoWMoM)*, Jun. 2013, pp. 1–6.
[11] L. Buttyán, T. Holczer, and I. Vajda, "On the effectiveness of changing pseudonyms to provide location privacy in VANETs," in *Proc. ESAS*, vol. 4572. Berlin, Germany: Springer, 2007, pp. 129–141.
[12] W. Gao, W. G. Hatcher, and W. Yu, "A survey of blockchain: Techniques, applications, and challenges," in *Proc. 27th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Jul./Aug. 2018, pp. 1–11.
[13] G. Danezis and S. Meiklejohn, "Centrally banked cryptocurrencies," 2015, *arXiv:1505.06895*. [Online]. Available: https://arxiv.org/abs/1505.06895
[14] G. Zyskind, O. Nathan, and A. Pentland, "Decentralizing privacy: Using blockchain to protect personal data," in *Proc. Secur. Privacy Workshops (SPW)*, May 2015, pp. 180–184.
[15] A. Dorri, M. Steger, S. S. Kanhere, and R. Jurdak, "BlockChain: A distributed solution to automotive security and privacy," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 119–125, Dec. 2017.
[16] L.-N. Lundbæk, D. J. Beutel, M. Huth, S. Jackson, L. Kirk, and R. Steiner, "Proof of kernel work: A democratic low-energy consensus for distributed access-control protocols," *Roy. Soc. Open Sci.*, vol. 5, no. 8, 2018, Art. no. 180422.
[17] P. Asuquo, H. Cruickshank, J. Morley, C. P. A. Ogah, A. Lei, W. Hathal, S. Bao, and Z. Sun, "Security and privacy in location-based services for vehicular and mobile communications: An overview, challenges, and countermeasures," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4778–4802, Dec. 2018.
[18] J. Xu, X. Tang, H. Hu, and J. Du, "Privacy-conscious location-based queries in mobile environments," *IEEE Trans. Parallel Distrib. Syst.*, vol. 21, no. 3, pp. 313–326, Mar. 2010.
[19] Z. Zhu and G. Cao, "Toward privacy preserving and collusion resistance in a location proof updating system," *IEEE Trans. Mobile Comput.*, vol. 12, no. 1, pp. 51–64, Jan. 2013.
[20] C.-Y. Chow and M. F. Mokbel, "Enabling private continuous queries for revealed user locations," in *Proc. Int. Symp. Spatial Temporal Databases*. Berlin, Germany: Springer, 2007, pp. 258–275.
[21] X. Pan, J. Xu, and X. Meng, "Protecting location privacy against location-dependent attacks in mobile services," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 8, pp. 1506–1519, Aug. 2012.
[22] J. Lin, W. Yu, N. Zhang, X. Yang, and L. Ge, "Data integrity attacks against dynamic route guidance in transportation-based cyber-physical systems: Modeling, analysis, and defense," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8738–8753, Sep. 2018.
[23] B. Palanisamy and L. Liu, "Attack-resilient mix-zones over road networks: Architecture and algorithms," *IEEE Trans. Mobile Comput.*, vol. 14, no. 3, pp. 495–508, Mar. 2015.
[24] R. Shokri, J. Freudiger, M. Jadliwala, and J.-P. Hubaux, "A distortion-based metric for location privacy," in *Proc. 8th ACM Workshop Privacy Electron. Soc.*, 2009, pp. 21–30.
[25] A. Lei, H. Cruickshank, Y. Cao, P. Asuquo, C. P. A. Ogah, and Z. Sun, "Blockchain-based dynamic key management for heterogeneous intelligent transportation systems," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 1832–1843, Dec. 2017.
[26] (2017). *Sawtooth: Proof of Elapsed Time*. [Online]. Available: https://sawtooth.hyperledger.org/
[27] L. Baird, "The swirlds hashgraph consensus algorithm: Fair, fast, byzantine fault tolerance," Hedera HashGraph, Tech. Rep. SWIRLDS-TR-2016-01, May 2016.
[28] M. Raya, P. Papadimitratos, and J.-P. Hubaux, "Securing vehicular communications," *IEEE Wireless Commun.*, vol. 13, no. 5, pp. 8–15, Oct. 2006.
[29] A. Festag, "Cooperative intelligent transport systems standards in Europe," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 166–172, Dec. 2014.
[30] *Dedicated Short Range Communications (DSRC) Message Set Dictionary*, Soc. Automot. Eng., DSRC Committee, 2009.
[31] *Intelligent Transport Systems(ITS); Security; Stage 3 Mapping for IEEE 1609.2*, document ETSI TS 102 867 v1.1.1, ETSI, Sophia Antipolis Cedex, France, 2012.
[32] S. Bao, W. Hathal, H. Cruickshank, Z. Sun, P. Asuquo, and A. Lei, "A lightweight authentication and privacy-preserving scheme for VANETs using TESLA and bloom filters," *ICT Express*, vol. 4, no. 4, pp. 221–227, Dec. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2405959517302333
[33] J. Freudiger, M. Raya, M. Félegyházi, P. Papadimitratos, and J.-P. Hubaux, "Mix-zones for location privacy in vehicular networks," in *Proc. ACM Workshop Wireless Netw. Intell. Transp. Syst. (WiN-ITS)*, 2007.
[34] Y. Pan, J. Li, L. Feng, and B. Xu, "An analytical model for random changing pseudonyms scheme in VANETs," in *Proc. Int. Conf. Netw. Comput. Inf. Secur.*, vol. 2, May 2011, pp. 141–145.
[35] R. Schlegel, C. Y. Chow, Q. Huang, and D. S. Wong, "User-defined privacy grid system for continuous location-based services," *IEEE Trans. Mobile Comput.*, vol. 14, no. 10, pp. 2158–2172, Oct. 2015.
[36] L.-N. Lundbæk, A. C. D'Iddio, and M. Huth, "Centrally governed blockchains: Optimizing security, cost, and availability," in *Models, Algorithms, Logics and Tools*. Cham, Switzerland: Springer, 2017, pp. 578–599.
[37] R. C. Merkle, "A digital signature based on a conventional encryption function," in *Advances in Cryptology—CRYPTO*. Berlin, Germany: Springer, 1987, pp. 369–378.
[38] C. Dwork and M. Naor, "Pricing via processing or combatting junk mail," in *Proc. Annu. Int. Cryptol. Conf.* Berlin, Germany: Springer, 1992, pp. 139–147.
[39] R. Yu, J. Kang, X. Huang, S. Xie, Y. Zhang, and S. Gjessing, "MixGroup: Accumulative pseudonym exchanging for location privacy enhancement in vehicular social networks," *IEEE Trans. Depend. Sec. Comput.*, vol. 13, no. 1, pp. 93–105, Jan. 2016.
[40] K. Emara, "Poster: PREXT: Privacy extension for veins VANET simulator," in *Proc. Veh. Netw. Conf. (VNC)*, Dec. 2016, pp. 1–2.
[41] D. Hankerson, A. J. Menezes, and S. Vanstone, *Guide to Elliptic Curve Cryptography*. Springer, 2006.
[42] W. Dai. (2009). *Crypto++ Library 5.6. 0*. [Online]. Available: http:/www.cryptopp.com

[43] *Intelligent Transport Systems (ITS); Security; Stage 3 Mapping for IEEE 1609.2*, Standard ETSI TS 102 867 v1.1.1, 2012.
[44] U. Rajput, F. Abbas, J. Wang, H. Eun, and H. Oh, "CACPPA: A cloud-assisted conditional privacy preserving authentication protocol for VANET," in *Proc. 16th IEEE/ACM Int. Symp. Cluster, Cloud Grid Comput. (CCGrid)*, May 2016, pp. 434–442.

SHIHAN BAO received the B.Eng. degree in communication engineering from the Nanjing University of Posts and Telecommunications, China, and the University of Northumbria, U.K., in 2014, and the M.Sc. degree from the University of Surrey, U.K., in 2015, where he is currently pursuing the Ph.D. degree with the Institute of Communication Systems. His research interests include privacy for vehicular networks and privacy protection for location-based services.

YUE CAO received the Ph.D. degree from the Institute for Communication Systems (ICS), University of Surrey, Guildford, U.K., in 2013. He is currently a Professor with the School of Transportation Science and Engineering, Beihang University, China. His research interest includes intelligent transport systems. He is an Associate Editor of the IEEE ACCESS, *EURASIP Journal on Wireless Communications and Networking* (Springer), *KSII Transactions on Internet and Information Systems*, and IGI Global *International Journal of Vehicular Telematics and Infotainment Systems*.

AO LEI received the B.Eng. degree from the Harbin Institute of Technology, China, and the University of Birmingham, U.K., in 2013, the M.Sc. degree from the University of York, U.K., in 2014, and the Ph.D. degree from the Institute of Communication Systems, University of Surrey, U.K., in 2017, where he was a Research Fellow. He is currently a Senior Engineer with Huawei Technologies, China. His research interests include security and privacy for vehicular networks and privacy protection for location-based services.

PHILIP ASUQUO received the B.Sc. degree in computer engineering from the University of Uyo, Nigeria, the M.Sc. degree in computer network technology from Northumbria University, U.K., and the Ph.D. degree from the Institute of Communication Systems, University of Surrey, U.K., in 2018. His research interests include cyber security of critical infrastructures, smart grid and smart homes, intelligent transport systems (ITSs), and wireless sensor network security.

HAITHAM CRUICKSHANK received the B.Sc. degree in electrical engineering from the University of Baghdad, Iraq, in 1980, the M.Sc. degree in telecommunications from the University of Surrey, U.K., and the Ph.D. degree in control systems from the Cranfield Institute of Technology, U.K., in 1995. He is currently a Senior Lecturer with the Institute of Communication Systems, University of Surrey. His research interests include network security and privacy, and satellite network architectures. He has involved in several European research projects in the ACTS, ESPRIT, TEN-TELECOM, and IST programmes. He is a member of the Satellite and Space Communications Committee of the IEEE Communications Society, and is also a Chartered Electrical Engineer and IEE corporate.

ZHILI SUN received the B.Sc. degree in mathematics from Nanjing University, China, and the Ph.D. degree from the Department of Computing, Lancaster University, U.K., in 1991. He is currently a Professor with the Institute of Communication Systems, University of Surrey, U.K. His research interests include wireless and sensor networks, satellite communications, mobile operating systems, traffic engineering, Internet protocols and architecture, quality of service, multicast, and security. He has been a Principal Investigator and a Technical Coordinator in a number of projects within the European Framework Program, including ESPRIT BISANTE, TEN-TELECOM, VIPTEN, GEOCAST, ICEBERGS, SATELIFE, and EuroNGI.

MICHAEL HUTH received the Ph.D. degree from Tulane University, New Orleans, LA, USA, in 1991. He has been a Professor with the Department of Computing, Imperial College London, since 2012. His research interests include cyber security, cryptography, mathematical modeling, and formal verification with applications in machine learning, FinTech, and the Internet of Things.