The Knowledge Engineering Review, Vol. 35, e3, 1 of 17.
doi:10.1017/S0269888920000028

Immutable autobiography of smart cars leveraging blockchain technology

MD. SADEK FERDOUS[1,2], MOHAMMAD JABED MORSHED CHOWDHURY[3], KAMANASHIS BISWAS[4,5], NIAZ CHOWDHURY[6], and VALLIPURAM MUTHUKKUMARASAMY[5]

[1] *Shahjalal University of Science and Technology, Sylhet, Bangladesh;*
e-mail: sadek-cse@sust.edu;
[2] *Imperial College London, London, UK;*
e-mail: s.ferdous@imperial.ac.uk;
[3] *La Trobe University, Melbourne, Victoria, Australia;*
e-mail: m.chowdhury@latrobe.edu.au;
[4] *Australian Catholic University, New South Wales, Australia;*
e-mail: kamanashis.biswas@acu.edu.au;
[5] *Griffith University, Queensland, Australia;*
e-mails: k.biswas@griffith.edu.au, v.muthu@griffith.edu.au;
[6] *Open University, Milton Keynes, UK;*
e-mail: niaz.chowdhury@open.ac.uk

Abstract

The popularity of smart cars is increasing around the world as they offer a wide range of services and conveniences. These smart cars are equipped with a variety of sensors generating a large amount of data, many of which are critical. Besides, there are multiple parties involved in the lifespan of a smart car, such as manufacturers, car owners, government agencies, and third-party service providers who also generate data about the vehicle. In addition to managing and sharing data among these entities in a secure and privacy-friendly way which is a great challenge itself, there exists a trust deficit about some types of data as they remain under the custody of the car owner (e.g. satellite navigation and mileage data) and can easily be manipulated. In this article, we propose a blockchain-assisted architecture enabling the owner of a smart car to create an immutable record of every data, called the *autobiography* of a car, generated within its lifespan. We also explain how the trust about this record is guaranteed by the *immutability* characteristic of the blockchain. Furthermore, the article describes how the proposed architecture enables a secure and privacy-preserving mechanism for sharing of smart car data among different parties.

1 Introduction

The popularity of smart cars (also known as connected cars) is increasing around the world as they offer a wide range of services and conveniences. A study carried out in 2018 estimated that the global connected car market was around 63 026 million USD with the prediction to reach around 225 158 million USD by 2025 (Jadhav & Sonpimple, 2018). This ever-increasing popularity of smart cars imposes a significant challenge from the viewpoint of the management and the security and privacy concerns of the data generated by such vehicles. The reason is manifold. Firstly, smart cars are embedded with a wide range of sensors which continuously generate data. For example, smart cars use sensors which provide real-time information about different components of the car in addition to traditional information such as total mileage. Secondly, multiple parties (e.g., manufacturers, government authorities, and service providers) are involved who generate data about the car throughout its lifespan and keep such records in their database.

Data generated by smart cars are extremely useful and, in some cases, are fundamental to continue its functionality. However, a major issue is that such data could be private in nature. Not only this, a smart car if connected with an analytic system, will enable the provision to collect an awful amount of data about its owners or whoever drives the car which ultimately can be exploited to generate and infer privacy-invasive knowledge. A few examples of such data and inferred knowledge are given below (International Transport Forum, 2019):

- **Contact information:** A smart car can know about the contact information of the owner when she syncs her smartphone with the infotainment system of the car.
- **Driving locations:** A smart car (with the help of the analytic system) can gather knowledge of the driving locations of the owner, her home and work addresses, favourite cafes/restaurants, or even the addresses of other family members.
- **Driving habit:** A smart car can infer the driving habit of its owner: how fast she drives, if she wears the seat-belt or not, breaking/acceleration habit of the driver, and so on.
- **Other information:** A smart car even can accumulate the music and shopping preferences of its owner, know their body weight using a weight-sensor underneath the driving seat, or social network information in case the owner logs into her social network accounts.

In addition to this, one major feature of smart cars is continuous internet connectivity, that is, such cars will be always connected to the Internet. This continuous connectivity introduces a novel threat which can be exploited by attackers with malicious intents. For example, in one of the most widely circulated smart car attacks reported in 2015, two security researchers named Charlie Miller and Chris Valasek demonstrated how an attacker could assume the full control of a smart car while the car was being driven on the highway (Greenberg, 2015; SmartCarIoT, 2017). They were able to switch on/off the music player of the car, control the driving wheel, switch on/off the AC, and more frighteningly, switch off the engine completely while the car was on the highway! Such incident could have disastrous ramifications on the safety of the car passengers in case the attacker has malicious intents.

Another important challenge is to ensure how data involving a smart car, both generated by itself or by other third parties, can be shared in an effective way. This problem, in fact, is like a two-edged sword. On the one hand, car owners lack any effective and proper mechanism to share smart car data with the third-party service providers (e.g. an insurer). On the other hand, third parties (e.g. a service station or a relevant Government agency) which generate data regarding a smart car also do not have a perfect solution which could enable the owner to share such data with another external party. The lack of such sharing mechanism essentially creates the 'silos' of information for any particular car. The effect of this is that there is no readily available global or holistic view of the information related to a specific vehicle.

In order to ensure the continuous growth of smart cars, these issues involving the management and sharing of such data must be addressed. All in all, users must be equipped with mechanisms which would allow them to create a holistic view of their smart car data with appropriate control and to share such data with anyone they intend to in a secure and privacy-friendly way. We argue that blockchain, a disruptive technology that has found many applications from cryptocurrencies to smart contracts, can be a potential solution to these challenges. In this article, we explore how a blockchain-supported architecture can be leveraged to address the identified issues.

Contributions. This article is an extension of our previous work presented at the 3rd Symposium on Distributed Ledger Technology, 2018 (Ferdous *et al.,* 2018). In that work, there have been the following contributions.

- We have presented a novel mathematical model to represent the holistic view of immutable records, called *autobiographies*, of smart cars.
- We have introduced a threat model to identify different security and privacy threats concerning a system involving smart cars.
- We have proposed a blockchain-based architecture which employs the concept of autobiographies of smart cars and can be utilized to allow different stakeholders to create a common platform so that

the owners of smart cars can share smart car data with other parties in a secure, transparent, and auditable way.
- We have qualitatively argued the resilience of the architecture against the security and privacy threats as identified by the threat model.

This article has been extended from the previous work with the following additions:

- We have expanded the background section with an enlarged discussion of blockchain, along with its properties and advantages, and smart cars.
- We have included a motivating scenario so as to exemplify different use-cases involving a smart car in daily lives which serve as the foundation for the protocol flows of the proposed system.
- We have extended the related work section significantly with additional discussion and critical analysis of a large number of relevant works.
- We have revised the protocol flows of the proposed architecture with new figures and further analysis.
- We have added a comparative analysis of our proposal with existing works which highlights its advantages over others.

All in all, with these additional contributions, the current version has been extended more than 40% from our previous work.

Structure. In Section 2, we provide a brief description about blockchain and smart car and its underlying issues. Section 4 provides a mathematical model to define what we mean by the term *autobiography* of a smart car. Section 5 presents a threat model, a requirement analysis, and the proposed system architecture along with a discussion of probable design issues. Finally, we conclude in Section 6 with directions for future work.

2 Background

Two disruptive technologies, *blockchain* and *smart car*, are among the most discussed subjects of recent time. These two apparently unrelated topics, however, share a close tie-up that can be capitalized on building the architecture for the next-generation applications to manage and share smart car data in a secure and privacy-friendly manner, what we have attempted in this article. A brief overview of both subjects would help to comprehend the remaining parts of the discussion well, and therefore, is presented next.

2.1 Blockchain

Bitcoin (Nakamoto, 2008), proposed in 2009, has emerged as the world's first widely used crypto-currency and paved the way for a technological revolution. It is underpinned by a clever combination of existing crypto mechanisms, which are now called blockchain technology or distributed ledger technology, providing its solid technical foundation. In recent years, blockchain has received widespread attention among the industry, the government, and academia. It is regarded as one of the fundamental technologies to revolutionize the landscapes of several application domains, by removing the need for a central trusted entity (Alexopoulos *et al*., 2017). At the centre of this technology is the blockchain itself which is a database consisting of consecutive blocks of transactions chained together following a strict set of rules. This database is then distributed and stored by the nodes in a peer-to-peer network where each new block of transactions is created and appended at a predefined interval in a decentralized fashion by means of a consensus algorithm. The consensus algorithm guarantees several data integrity-related properties (discussed below) in the blockchain. The term blockchain is often synonymized with another term *distributed ledger*. However, we differentiate between these two terms in the sense that distributed ledger is a more generic term. A blockchain is just an example of a distributed ledger whereas there could be other types of distributed ledger.

Evolving from the Bitcoin blockchain, a new breed of blockchain platforms has emerged which facilitates the deployment and execution of computer programs, known as smart contracts, on top of the respective blockchain. Such smart contracts enable the creation of so-called decentralized applications (DApps), which are autonomous programs operating without relying on any human intervention. Being part of the blockchain, smart contracts and their executions become immutable and irreversible, a sought-after property having a wide range of applications in different domains. One prominent example of such a blockchain platform is Ethereum (Dannen, 2017; Ethereum, 2018) which is often regarded as *Blockchain 2.0*.

Some of the major characteristics of blockchain platforms are (Ferdous *et al*., 2019):

- Distributed consensus on the blockchain state.
- Immutability and irreversibility of the blockchain state.
- Data provenance of transactions guaranteed by cryptographic mechanisms.
- Accountability and transparency of blockchain data and actions.
- Decentralization within the system providing strong resiliency amid system failures and attacks.

Equipped with these characteristics, blockchain platforms offer significant advantages over traditional systems for many application domains. However, depending on the application domains, different blockchain deployment strategies can be pursued. Based on these strategies, there are two predominant blockchain types: public and private. A public blockchain, also sometimes called as the *unpermissioned blockchain*, allows anyone to join and create and validate blocks as well as to modify the blockchain state by storing and updating data by means of transactions among participating entities. On the other hand, a private blockchain, also may be called as *permissioned blockchain*, will only allow authorized entities to join and participate in blockchain activities with the aim to ensure some form of accountability for the transaction of data, which might be desirable in some use-cases.

2.2 Smart car

A smart car is a car which is equipped with a system that collects real-time information about its different components and processes these information to provide value-added operation and maintenance features to the car owner (ENISA, 2016). It improves the car owner experience and the car safety. It allows the communication with other smart cars and provides different types of telemetries. All in all, the integration of almost every aspects of our lives with continuous internet connectivity is driving the popularities of smart cars.

As stated earlier, a smart car bundled with sensors generates different types of data. In addition, there might a number of systems which either utilize such data for providing different services to the its users or accumulate other data by interacting with its users (Coppola & Morisio, 2016). Examples of such services are, but not limited to, infotainment services allowing users to connect to their different smartphone apps, intelligent transport services, and other convenient services. Understandably, some of the corresponding data are critical in nature while some other are private. Therefore, a system to support data management and sharing such smart car data must consider different security and privacy issues. Among them, the major issues are highlighted below:

- **Privacy:** Many of the generated data are sensitive in nature.
- **Security:** Apart from the security of the crucial internal components of a smart car, the data generated by smart cars lack any proper mechanism to guarantee the confidentiality, authenticity, and integrity.
- **Holistic view:** The size of data generated by different sensors of a smart car can be huge. Even worse is that there are third parties, e.g. service centres, smart infrastructures, and so on, who might have data about the smart car. The scattered nature of such data makes it difficult for any owner to have a holistic view during the lifespan of a smart car.
- **Sharing:** The traditional approach makes it difficult to share smart car data in a secure, privacy-friendly, and auditable way among different parties.

3 Related work

Over the last few years, researchers have published different research works related to security and privacy issues in smart or connected cars. For examples, authors in Fraga-Lamas & Fernández-Caramés (2019) and Halder *et al*. (2019) have reviewed different issues, including security and privacy, concerning the integration of blockchain technology for automotive industry in general.

In a more application-oriented work, Rouf *et al*. have presented different security and privacy vulnerabilities in the wireless network of the smart car (Roufa *et al*., 2010). In particular, they have experimented with wireless tire pressure monitoring systems and have been to able to retrieve the 32 bit identifier of the respective car by remote eavesdropping and then reverse engineering. Even more, they have managed to remotely trigger a false tire pressure warning message in a moving vehicle from another vehicle.

Woo *et al*. have demonstrated a long-range wireless attack using a real vehicle and malicious smartphone application in a connected car environment in Woo *et al*. (2015). The attack targets the in-vehicle network called Controller Area Network which is utilized by the electronic control units (ECUs) within a smart car. To mitigate this attack, they have also proposed a security protocol which they claim to be more efficient and secure than the existing ones.

Dorri *et al*. have proposed a blockchain-based system to overcome the security and privacy issues in smart cars (Dorri *et al*., 2017). They have demonstrated the applicability of their system by exploring two use-cases: remote software updates and dynamic vehicle insurance fees using blockchain. However, they have not provided any mechanism to share car-related data with other interested parties or how to build a holistic view of smart car data.

In (2018), Cebe *et al*. have proposed a blockchain-based framework for storing various data from numerous sensors of any modern vehicle in order to facilitate a forensic analysis should any accident happens. Such data are crucial in post-accident scenarios for identifying the faulty parties and hence, as they have argued, there is a need to store such data in a tamper-proof manner. Their proposed blockchain-enabled framework is envisioned in such a way so that such data stored within the blockchain can be shared easily using the corresponding blockchain protocol with respective parties such as car owners, manufacturers service centres, government agencies, and insurance companies. This enables these parties to retrieve the corresponding data for a vehicle in case a forensic analysis is required. Being stored in the blockchain with its corresponding hashes, the integrity of such data is guaranteed which eliminates the need for any trusted party during the forensic analysis. This work shares a similar vision to our proposal: specifically, with respect to the data sharing utilizing a blockchain-based infrastructure. However, their proposal is restrictive in nature in the sense that they have only focused on a specific application (forensic application) with a particular set of sensor data suitable for such application, without considering other types of sensor data. In addition, their proposal to store all sensor data in the blockchain might be infeasible considering current blockchain systems are inappropriate to store a massive amount of data in the blockchain due to associated cost, scalability, and blockchain bloating issues (Dumont, 2019). Also, they have not considered how to build a holistic view of all smart car data.

The work presented in Sharma *et al*. (2018) is quite relevant to the scope of this article. There, the authors have outlined a blockchain-enabled framework for automotive industry considering different actors such as manufacturers, dealers, and regulators. They have briefly and vaguely sketched out the mechanisms required at different stages of the life cycle within the automotive industry. They have also presented a novel algorithm which utilizes a fruit fly optimization algorithm (FOA) for miner node selection in the underlying blockchain network. Even though their proposal is relevant and similar in scope to the proposal presented in this article, we highlight a few inconsistencies in their work and major differences with our work. Their proposal has utilized Ethereum within a private network where different agencies are assumed to participate as miner nodes. We believe that there are better alternatives to private Ethereum for creating a private blockchain network. For example, Hyperledger Fabric (Fabric, 2019) is an enterprise-grade private blockchain platform that does not require any mining node to participate in any mining procedure. As such, there is no need to utilize their FOA algorithm. It is also unclear if they have considered the need for privacy in their framework. Utilizing a private Ethereum would allow everyone in the network to get access to every piece of data stored in the blockchain, which would seriously

undermine the privacy of any sensitive data. Furthermore, the interactions among different organizations have not been detailed out. For example, they have not provided the protocol flow by which different entities will access any car data by strictly following any access control mechanism. Finally, they have not considered the scenario for creating a holistic view of car data and sharing it with other entities in the framework.

The authors in Singh & Kim (2017) have argued that intelligent vehicles require a trusted environment in order to facilitate a fast, reliable, and secure transmission of data among themselves. To address this challenge, they have proposed a blockchain-supported data sharing framework to disseminate different types of data of an intelligent vehicle (e.g. self-driving cars) to other nearby vehicles in a secure, reliable, and trusted way. Unfortunately, their proposal is quite vague and without any critical analysis of a threat model for the underlying application. They also did not consider how such data can be shared with other parties nor they did consider the need for creating any holistic view of data generated within a vehicle.

With the advent of autonomous vehicles, it is expected that a plethora of Machine to Machine (M2M) interactions with automatic transactions will commence. With this argument, the authors in Pedrosa & Pau (2018) have proposed to utilize a smart contract-based payment system (e.g. Ethereum) for automatic contract execution and transaction. With specific focus on vehicle-to-charging station use-case scenarios, several algorithms have been presented to ensure an immutable, flexible, and scalable M2M transactions using Ethereum. With a different use-case and scope, this article is not relevant to the scope of the current article.

There have been several proposals to utilize blockchain technology for maintaining Vehicle to Vechicle (V2V) and vehicle-to-infrastructure (V2I) communications involving intelligent vehicles such as smart cars as presented in Singh & Kim (2018), Shrestha *et al*. (2018, 2019), Bartolomeu & Ferreira (2019), Lei *et al*. (2019). Since these proposals are not relevant to the scope of this article and hence, they have not been explored further.

Conversely, a few proposals are more authentication-oriented, in the sense they aim to improve the authentication mechanism within the domain of internet vehicles. For example, authors in Sharma & Chakraborty (2018a) have proposed mechanisms to securely authenticate vehicles and enforce a decentralized access control mechanism using a blockchain-based infrastructure as well as to preserve the privacy of the communicating vehicles while they communicate. In a similar vein, Wang *et al*. have put forward the idea of employing blockchain systems in order to carry out secure identity authentication for smart vehicles (Wang *et al*. 2019). Even though both these proposals cover the domain relevant to this article, they focus on a very specific application and do not cover the whole spectrum of application scenarios we consider in our proposal.

There have been a few application-oriented works as well. For example, authors in Cintron *et al*. (2019) have explored the possibility of utilizing blockchain technology for secure event attestation. The authors argue that a vehicular network would likely be in an open network where any vehicle can participate. In such, a malicious vehicle can disrupt the flow of the network by generating and propagating falsified event messages. The proposal presented in Cintron *et al*. (2019) is aimed to tackle this problem. However, their focus is not properly aligned with that of this article. Similarly, a blockchain-based proposal has been presented in Sharma & Chakraborty (2018b) in which the authors have introduced the idea of a blockchain-based infrastructure for managing data generated by vehicles in a vehicular network. In particular, the proposed infrastructure exploits blockchain for a secure access control as well as a reliable dissemination of messages generated within the vehicular network.

4 Mathematical model

In this section, we present our idea of how we can conceptually represent the holistic view of a smart car data using a novel concept called the *autobiography* of a smart car. To concretize the semantic of the concept, we also build a mathematical model.

As per the ENISA report (ENISA, 2016), a smart car has the following components where each component has multiple sub-components/sensors.

- *Powertrain control:* Engine control, transmission control, speed and gear control, power train sensors, and so on.
- *Chassis control:* Steering control, braking system, Advanced Driver Assistance System (ADAS), and so on.
- *Body control:* Door locking, air conditioning sensor, light sensors, seat-belt sensor, and so on.
- *Infotainment control:* Audio-visual unit, navigation, external media unit, and so on.
- *Communications control:* Gateway ECUs, telematics, communication unit, and so on.
- *Diagnostic and maintenance systems:* On-board Diagnostics (OBD) ports, external aftermarket dongles, and so on.

Each of these components/sensors can generate data. Since these components are internal as far as a smart car is concerned, we call them *internal sources*. However, there could be data sources which are *external*, such as a service centre providing a yearly fitness certificate, Government transport authority providing a yearly tax certificate as well as registration certificates, and even other vehicles (enabling V2V communication) or infrastructures (enabling V2I, Vehicle to Infrastructure, communication).

Now, we gradually build the mathematical model by denoting the set of smart cars with VEC. We develop the mathematical model from the perspective of a smart car $vec \in VEC$. The set of other cars is presented using the notation VEC', such that $VEC' \triangleq \{VEC \setminus vec\}$.

We use the notation S_{vec} to denote the set of internal sources of the smart car vec. The external sources, as we consider within the scope of this article, consist of other vehicles (belonging to the set VEC'), set of infrastructures (denoted with INF), and set of organizations (denoted with O). Service centres or officials regulatory bodies are examples of different organizations within the scope of this article. Therefore, the set of external sources denoted with S_e is defined as: $S_e \triangleq \{VEC' \cup INF \cup O\}$.

Then, we can define S^{vec} (the set of all sources) from the perspective of the smart car vec as the union of its internal and external sources. Thus,

$$S^{vec} \triangleq \{S_{vec} \cup S_e\}$$

Based on this foundation, we define the concepts of claim and assertion from the perspective of a smart car:

DEFINITION 1 ***Claim:*** *A claim is a statement about a smart car consisting of information, relevant to that smart car, generated either by one of the internal sources or by an external source. The information, thus, can be anything generated by a single sensor. Alternatively, this information can also be generated by another vehicle, by an infrastructure (e.g. traffic light), or by any regulatory or authorized entity.*

DEFINITION 2 ***Assertion:*** *An assertion is a signed collection of claims. We differentiate between two types of assertions: internal or external. An internal assertion is the collection of claims where each claim is generated by one of the internal sources of the car and signed by the private key of the smart car. On the other hand, an external assertion is a collection of claims produced by one of the external sources and subsequently signed by the respective private key of the source.*

Practically, a claim can be represented using a structured data consisting of name-value pairs where a name presents a property and the value represents its corresponding information. This concept of structured data can be utilized to formalize the notion of a claim. Let C^s_{vec} denote the set of claims for a car vec generated by an internal source $s \in S_{vec}$. Then, a claim $c \in C^s_{vec}$ can be defined in the following way:

$$c \triangleq \langle\ (n_1, v_1), (n_2, v_2), (n_3, v_3)...(n_j, v_j)\ \rangle$$

Here, n represents the name of the property, v presents its value, and $j \in \mathbb{N}$.

Similarly, we can denote the set of claims for a car vec generated by an external source $s' \in S_e$ using the notation: $C^{s'}_{vec}$ where each single claim consists of name-value pairs as defined before for an external source. A visual illustration of claims is presented in Figure 1.

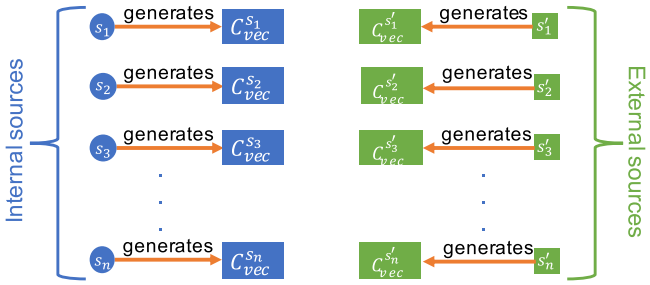

Figure 1 Claims from different sources

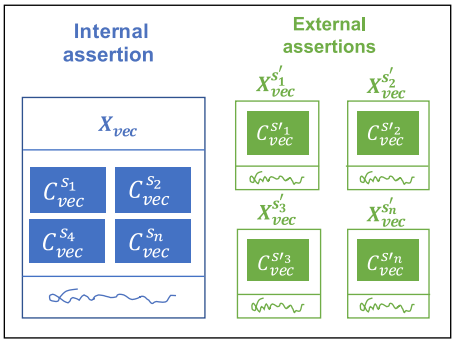

Figure 2 Modelled autobiography of a smart car

An internal assertion, denoted as X_{vec} for a car *vec*, can be modelled as a set containing the signed union of all claims generated by the internal sources of a car and defined it in the following way:

$$X_{vec} \triangleq \left\langle \left\{ \bigcup_{s \in S_{vec}} C^{s}_{vec} \right\}_{K^{-1}_{vec}} \right\rangle$$

Here, K^{-1}_{vec} represents the private key of the vehicle *vec*.

Finally, we can define an assertion provided by an external source ($s' \in S_e$) regarding a vehicle *vec* in the following way:

$$X^{s'}_{vec} \triangleq \left\langle \left\{ C^{s'}_{vec} \right\}_{K^{-1}_{s'}} \right\rangle$$

Here, $K^{-1}_{s'}$ represents the private key of the external source s'.

Next, we provide a definition of the autobiography of a smart car:

DEFINITION 3 ***Autobiography:*** *The autobiography of a smart car is defined as the collections of internal and external assertions generated throughout its lifespan.*

Denoted with BIO_{vec} to represent the autobiography of a smart car *vec*, we define it in the following way:

$$BIO_{vec} \triangleq \left\langle X_{vec} \bigcup \left\{ \bigcup_{s' \in S_e} X^{s'}_{vec} \right\} \right\rangle$$

This definition is visually presented in Figure 2.

According to this definition of the autobiography, it consists of every piece of information generated by every single internal and external source over the lifespan of a car. Understandably, this can be a large collection of information and might be unsuitable or even not required for exchange and

sharing with parties. For such scenarios, we introduce a novel terminology called *snapshot*, defined in the following way:

DEFINITION 4 ***Snapshot:*** *The snapshot of a smart car (denoted with $SNAP_{vec}$ for a smart car vec) is the collections of internal and external assertions generated with only the required information from the respective sources.*

Essentially, it is created with the subset of information generated and signed by the respective source. Intuitively, it can be modelled in the following way:

$$SNAP_{vec} \triangleq \left\langle \mathcal{X}_{vec} \bigcup \left\{ \bigcup_{s' \in S_e} \mathcal{X}_{vec}^{s'} \right\} \right\rangle$$

Here, $\mathcal{X}_{vec} \subseteq X_{vec}$ and $\mathcal{X}_{vec}^{s'} \subseteq X_{vec}^{s'}$.

5 The proposed system

In this section, we present the proposed system with the sole purpose to tackle the identified issues. We start with motivating scenarios (Section 5.1) and then follow with a threat model (Section 5.2.) and requirement analysis to mitigate such threats (Section 5.3), then propose an architecture (Section 5.4), and finally analyze the architecture (Section 5.5).

5.1 Motivating scenario

In this section, we present a motivating scenario which exemplifies different use-cases involving a smart car in daily lives. We will utilize these use-cases as the basis to model threats, formulate requirements, and design our architecture.

Let us suppose Mrs Alice has bought a smart car for her daily use. The vehicle has all the required sensors and facilities as standard in a smart car. It is also equipped with cellular connectivity (e.g. 4G network connection) so that it can remain connected to the Internet via a mobile network and a hefty internal storage device to store car data. As she uses the car, the sensors generate data that the car stores in its internal storage. The car also has an interface by which it can be connected to an external device (e.g. PC/laptop) to download the stored car data into the device. There is an app, provided by the car vendor, which can connect to the smart car and aggregate car data in a meaningful fashion (e.g. total accumulated mileage) and display it on the app.

During the vehicle's lifetime, there is a maintenance requirement that it needs to be taken to a service centre for servicing. Once the servicing is finished, its record is stored in the car as well as into the system of the servicing centre. Also, the car needs to be tested for fitness each year in a service centre once the car is certain years old (e.g. 3 years) and taxed per the law of the vehicle regulatory authority of the country where Alice resides. The fitness and tax certificates are stored in the car as well as in the system of authorities.

Alice also needs to insure the car as per the law and is provided with an insurance certificate. To accurately calculate the correct premium, it will be beneficial for the insurer to have a proper knowledge regarding the fitness certificate as well as the service data. Alice can retrieve this information from her car and share it with the insurer. Once an insurance certificate is provided, it can also be stored in the car. We also assume another scenario when Alice is involved in an accident, and the car needs to be repaired in a service centre. The repair record is stored in the car and shared with the insurer for the insurance claim. Such insurance claim data are also stored in the car.

Finally, we consider the scenario when Alice sells her car to another buyer. To ensure her privacy, she can delete any fine-grained data stored in the car. However, some aggregated data (e.g. total mileage) are not allowed to be removed from the car storage in order to ensure the authenticity of data. Also, Alice is

not allowed to remove any car certificate (fitness/tax) or data provided by a third party in order for the buyer to check the correct status of the car. It is to be noted that we consider the insurance certificate is a private data and in such, she is allowed to remove insurance data from the car storage.

5.2 Threat model

Threat modelling is one of the mostly used methods to identify, communicate, and understand threats and mitigation mechanisms within the context of protecting (IT) assets, smart car in the context of this article. However, it is highly dependent on the capabilities of the attacker that is assumed for a particular system. Traditionally, the capabilities of attackers are presented using an adversary model. One of the most well-known adversary models is the honest-but-curious model (Paverd *et al.*, 2014) which assumes that an attacker participates in the protocol of the system and behaves correctly. She can intercept, send (or receive) any message to (or from) the protocol to which it participates. However, she cannot modify any message or launch other types of attacks. We believe this is quite restrictive in the sense that modern attackers can exhibit additional capabilities.

We also differentiate between two types of attackers, namely *internal attacker* who is the owner of the car motivated to gain illicit advantages by fabricating sensor-generated data (e.g. odometer) and *external attacker* who can be either an individual other than the owner or an organization whose main motivation is to act maliciously by getting hold of data from the smart car or disrupting its system or other external systems which might deal with the autobiography of the smart car. With this conjecture in mind, we also assume that an attacker can (i) intercept as well as modify any network packets transmitted over an insecure communication channel, (ii) disrupt the internal system of the smart car or other external systems which deal with the autobiography of the smart car, (iii) gain unauthorized access only if she has access to the required credential, e.g. username/password pair or secret key, and (iv) try to alter data and certificates generated by different sensors stored within a smart car or provided by external entity, respectively, which ultimately will alter the autobiography of a smart car.

Considering the above capabilities, we have chosen a well-established threat model called STRIDE (Shostack, 2014) developed by Microsoft. The STRIDE model is briefly presented below.

- **T1-Spoofing Identity:** The act of spoofing refers to an adversary using the identity of an authorized user (e.g. owner) to illegally access or modify resources on a system where they would normally do not have any access.
- **T2-Tampering with Data:** As many of the car data, including the certificates provided by external parties, remain under the control of the users, they can try to change the value of the data (e.g., odometer reading) or fitness/servicing record. Therefore, the integrity of the data can be lost.
- **T3-Repudiation:** This involves a user or attacker who leverages the inability of a system to track invalid and illegal actions and uses them to gain some advantages in the system.
- **T4-Information Disclosure:** Private or sensitive data stored in a smart car may be leaked when the car is sold to another user.
- **T5-Denial of Service (DoS):** The system that is used to access the smart car data or to share the autobiography can be the target of a DoS attack.
- **T6-Elevation of Privilege:** Malicious software with potential exploitable vulnerabilities may be the first step to gain access to external systems (e.g. systems utilized by service centres or other Trusted Third Parties, TTPs), thus potentially gaining a privileged access on a large set of vehicles.

Among all these threats, we exclude the elevation of privilege from our consideration for this article. This is because such a threat is more relevant for enterprise systems. Besides, with the involvement of very sensitive private data handled by smart cars, we must also consider the privacy threats related to the lack of control, in addition to the standard STRIDE model.

- **T7-Lack of Control:** The owner has no or little control regarding how data of the smart car are shared with other entities.

Table 1 Inter-relation between requirements and threats

Threats	**Requirement(s)**
T1	*S3*
T2	*S1*, *S2*, *S3*
T3	*S1*, *S2*, *S3*
T4	*S1*, *S4*
T5	*S6*
T6	Not considered
T7	*P1*, *P2*

5.3 Requirement analysis

In this section, we present a set of security and privacy requirements which can be utilized to mitigate the identified threats. The security requirements are:

- **S1:** The data generated in a smart car must be stored in a secure way with a guarantee of its confidentiality and integrity.
- **S2:** Data relevant to the smart car must be transmitted over a secure channel.
- **S3:** There must be a secure way to access the data and create an autobiography with strong guarantee of its integrity. This mitigates *T1*.
- **S4:** The sharing of the car data must be carried out in a secure, transparent, and accountable manner. *S4* in combination with *S1* can tackle *T4*.
- **S5:** The authenticity and non-repudiation of data must be guaranteed. *S5* in combination with *S1*, *S2*, and *S3* can mitigate *T2* and *T3*.
- **S6:** A distributed system should be leveraged in order to minimize the impact of any DoS attack on the relevant system. This tackles *T5*.

Next, the privacy requirements are presented:

- **P1:** The sharing platform must satisfy the selective disclosure property, allowing the owner to build an autobiography with the full control over the data.
- **P2:** The autobiography must be shared with the explicit consent of the owner. *P1* and *P2* combinedly can mitigate *T7*.

In Table 1, we present the inter-relation between the requirements and the threats they can mitigate.

5.4 Architecture

The architecture of the proposed platform is presented in Figure 3. A brief overview of different aspects of the architecture is provided next with an illustration of how this architecture can be utilized for different use-cases.

Organization and Setup. The architecture is centred around a smart contract-supported blockchain platform. Different nodes connected to the blockchain network are the smart car vendors, smart cars, different government regulatory agencies, car insurers, and other third-party service providers such as service centres. There is a setup phase at the initial stage which is illustrated in Figure 4. The car vendor during the production of the car (denoted as $vec \in VEC$) will generate a key pair (public key K_{vec} and private key K_{vec}^{-1}) utilizing the blockchain platform. The private key is stored in a tamper-proof hardware of the smart car while the public key is included as a digital certificate signed by the vendor.

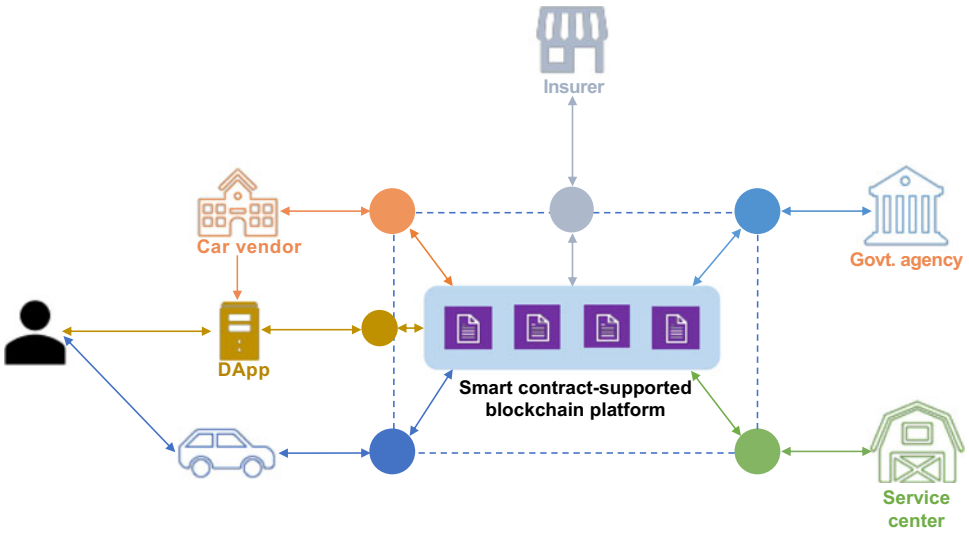

Figure 3 Architecture of the proposed platform

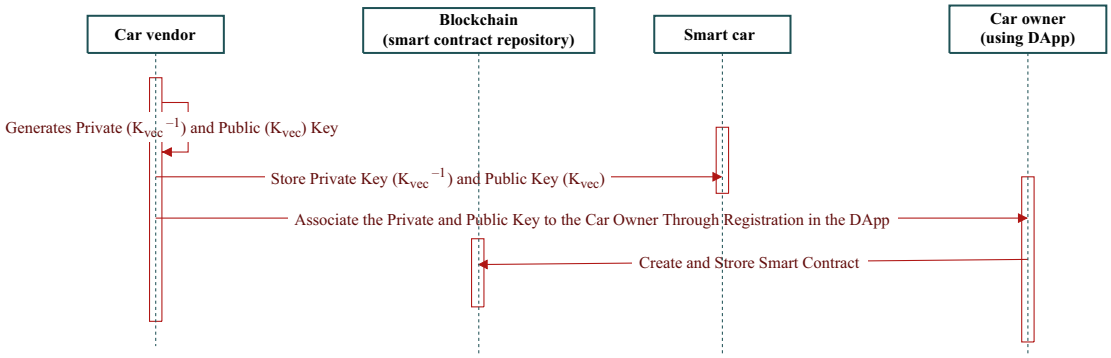

Figure 4 Initial setup

The car vendor provides a decentralized application (the so-called *DApp*) to interface with the car and the blockchain platform. The owner can utilize a web, mobile, or desktop application, all provided by the vendor, to interact with the DApp. When a user first buys the car, she needs to go through an initialization phase in which she needs to deploy the smart contract via a transaction, signed by the private key of the car, using the web, mobile or desktop app, and DApp. This smart contract will be utilized to store or retrieve information (mostly the hashes) regarding the respective smart car. Intuitively, separate smart contracts will be required for separate cars. In addition, other nodes will also utilize their own smart contract to facilitate different functionalities.

We also assume a special type of smart contract, called *repository*, that holds the identity of the car, its corresponding public key, and the address of the smart contract for all relevant entities within the network. In such, this smart contract acts as the one-stop entry to retrieve the identity-public key pair or the identity-smart contract address pair for a particular entity in the network. This repository contract can be maintained by the vendor itself or by any third party.

Claim management. A smart car generates different claims in its lifespan. However, for simplicity, we make the assumptions that (i) a smart car is always connected to the network, (ii) a claim is generated in a specified duration and is stored in the internal storage of the smart car and its hash is stored in the blockchain via a transaction signed by the private key of the car, and (iii) an internal assertion is created when a trip is completed where the assertion is stored in the internal storage of the car and its hash is stored in the blockchain via a transaction which is signed by the private key of the car.

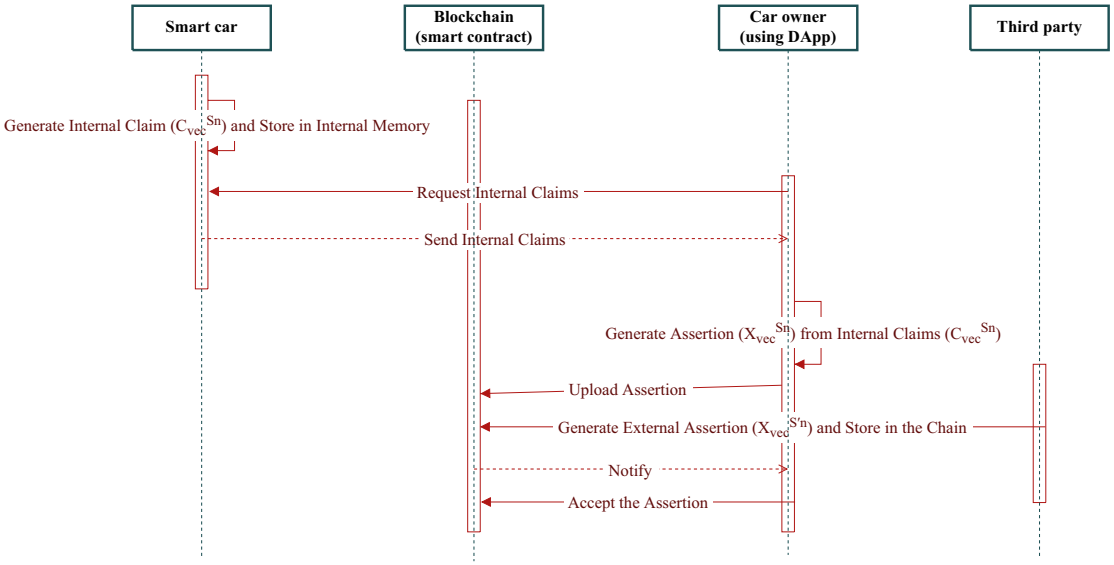

Figure 5 Creation and storage of assertions

The owner of a smart car can always connect to the smart car to download different claims and assertions either via the web, mobile or desktop app and store them securely in encrypted formats in an external storage of the owner. Whenever such data are downloaded from the smart car, the integrity of each data is checked with its corresponding hash from the blockchain. If the hashes do not match, it implies that the data stored in the smart car have been altered and hence, an error message is displayed to the user and a marker is set to signify this.

Assertion creation. One of the core functionalities is the creation of assertions, e.g. a tax certificate and a fitness certificate, by the owner and by other third parties. We envision the following way for the creation of internal assertions by the owner and transfer of external assertions from a third party to the owner of a smart car. The assertion creation process is illustrated in Figure 5.

The owner can utilize her preferred app and the DApp to create an internal assertion any time she likes. When this process is started, individual claims are aggregated into an internal assertion which is then digitally signed by the private key of the car and stored in the smart contract.

When an external third party would like to share an external assertion, it can retrieve required information, public key, and address of the respective smart contract, by just looking up in the repository smart contract. Then, the third party can just create an assertion, optionally encrypted it with the public key of the car and then send over the respective smart contract of the car via a signed transaction. Once the car smart contract receives the transaction, it notifies the owner and upon being approved, the assertion is then stored in the smart contract, which later can be retrieved using any preferred device.

An exemplary scenario of this use-case when a user is involved with an accident with her smart car and thus, requires repairing. However, being insured, the user would like to reclaim the repairing bill from the insurer. The flow in such situation is as follows. The car is taken to a service centre where the repair is carried out. Once completed, the service centre generates an assertion, optionally encrypted, stating the type of the repair and the required bill. The assertion is then stored into the smart contract as discussed previously. The user, then, interacts with the smart contract of the insurer to share her assertion with the insurer. The insurer, upon verification, reimburses the repair bill and issues an insurance bill certificate. This bill certificate is then stored, as an assertion, into the smart contract of the smart car using the flow discussed previously.

Autobiography creation. The owner can use an app to combine different internal and external assertions (X_{vec} and $X_{vec}^{S'}$) into an autobiography (BIO_{vec}) of the smart car with its hash stored in the blockchain. Eventually, when the autobiography is re-updated, the hash is also updated.

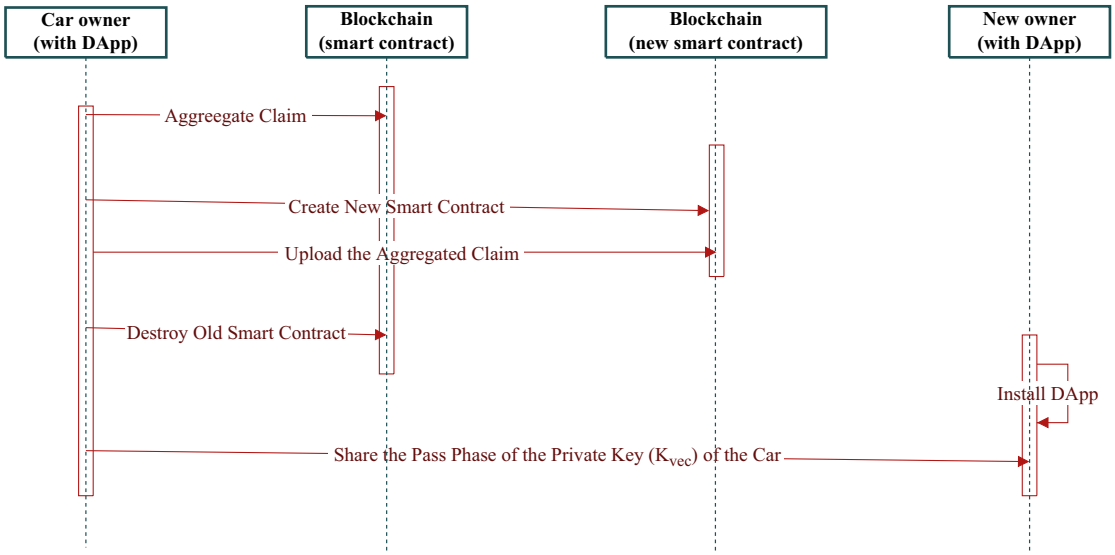

Figure 6 Change of ownership

The size of such an autobiography will be a significant issue when data from different sensors are combined. Additionally, having the whole spectrum of data into the autobiography will raise different privacy questions. That is why it might not be ideal to share it with an external entity (e.g. an insurer). Towards that aim, a user can choose the information that is to be shared to create a snapshot of the autobiography ($SNAP_{vec}$). The owner then looks into the repository smart contract to retrieve the public key and smart contract address of the external entity and sends the snapshot to the smart contract. The snapshot can be encrypted for additional security using the public key of the external entity. Once the smart contract of the external entity receives the snapshot, it goes through the same verification process as described before.

Change of ownership. When the ownership of a smart car changes, the architecture proposes a special type of ownership transaction. This will be initiated by the new owner and verified by a smart contract of a Government agency. The ownership change process, illustrated in Figure 6, will involve the following steps.

- Create an assertion from an aggregated version of all claims from the car as generated while in the possession of the previous owner in a privacy-friendly way. This is to ensure that the new owner cannot get hold of sensitive private data (e.g. driving locations) of the old owner while preserving the essence of the usage of the car. One example of this is that the new assertion contains the total mileage of the smart car under the old owner instead of fine-grained details of where, when, and for how long the car was driven previously.
- Import all data (mainly assertions) from the corresponding smart contract under the old owner and create a new smart contract under the new owner.
- Destroy the old smart contract in such a way that there is no way to retrieve data from it.

These steps will ensure that the important information regarding the smart car does not get lost with the change of ownership while preserving the privacy of the old owner.

5.5 Analysis and design choice

In this section, we analyze the ways the proposed architecture can satisfy different requirements. Data generated by different sensors of the smart car are stored, as per the proposal, either in a tamper-proof hardware while in the car storage or in encrypted format when downloaded in an external storage

Table 2. Comparison of other works with our proposal

	Sec. Req.						**Priv. Req.**		**Other features**	
	S1	***S2***	***S3***	***S4***	***S5***	***S6***	***P1***	***P2***	**Holistic view**	**Autobio.**
Dorri *et al*. (2017)	√	√	X	X	X	√	X	X	X	X
Cebe *et al*. (2018)	√	X	X	X	√	X	X	X	X	X
Sharma *et al*. (2018)	√	X	X	√	X	√	X	X	X	X
Singh and Kim (2017)	X	X	X	√	X	X	X	X	X	X
Our proposal	√	√	√	√	√	√	√	√	√	√

or stored in the different smart contracts in the form of assertions, autobiography, or snapshot. In addition, their hashes are stored in the blockchain, guaranteeing its integrity. All these satisfy the *S1* requirement.

The protocol for the architecture can be deployed in such a way that data always are transmitted over an encrypted channel, thereby satisfying *S2*. To ensure *S3*, the desktop, web, or mobile app should be equipped with a strong authentication mechanism so that only the owner, or an authorized delegated entity, can access as well as generate the autobiography or the snapshot involving the car data. The proposal requires that the hashes of the assertions as well as autobiography and snapshot are stored in the blockchain. This creates an immutable evidence for the corresponding element which satisfies *S3*.

The sharing of assertions and snapshots is always carried out within the blockchain via signed transactions. This creates transparent and auditable trails while ensuring the security and thus satisfies *S4*. The signed assertions and transactions provide a strong guarantee of authenticity and non-repudiation as it is assumed only the authorized owner can utilize the corresponding private keys. This satisfies *S5*. An added advantage of a blockchain-supported platform is its decentralization capabilities. This minimizes the impact of any DoS attack and hence satisfies *S6*.

The proposed system advocates that each internal assertion and the snapshot is created with the involvement and explicit consent of the owner with the ultimate authority to decide what information is to be included in such elements. This satisfies *P1* and *P2*.

Next, a comparative analysis of our proposal with the existing relevant works is presented in Table 2. It is to be noted that the comparison is based on different security and privacy requirements and other additional features. In this table, *Sec. Req.* is the shorter representation of *security requirements,* whereas *Priv. Req.* and *Autobio* imply *privacy requirements* and autobiography, respectively. Moreover, we have the '√' symbol to denote a particular requirement/property that has been considered by the corresponding work, whereas 'X' signifies that the requirement/property has not been considered.

As evident from Table 2, none of the works has explored the idea of creating a holistic view of smart car data in order to ensure their secure sharing among trusted parties. Since the idea of creating autobiographies of smart cars is novel in nature, it is understandable that it has not been explored in any previous work. However, the lack of any concrete privacy-preserving consideration in the existing works for sharing smart car data (or intelligent vehicles data in general) is puzzling and worrisome. In addition to this, our analysis suggests that many existing works have not considered several crucial security requirements. In comparison to this, our proposed system has addressed all these gaps.

To realize such a proposal, there are several design choices that need to be finalized. Firstly, the type of a blockchain platform that needs to be utilized for the proposal. We envision a private blockchain platform such as Hyperledger Fabric (Fabric, 2009). This is because a private blockchain offers a better support of privacy, scalability, and throughput. Secondly, the next issue is the responsibility to maintain such a blockchain. We believe a consortium of car vendors, Government bodies, and third parties, as outlined in our article, can be a good choice. This will ensure a shared responsibility between the inter-connected parties and greater benefits for everyone involved within this ecosystem.

6 Conclusion

Privacy and security have been becoming more critical as modern cars are getting smarter. The smart cars themselves and participating organizations generate data about the vehicle that remain under the custody of different parties including the car owner. Building trust about the data, securing storage, and sharing data in a privacy-preserving manner are among the main future challenges. In this article, we have proposed a system architecture that can address these three challenges using blockchain technologies and a novel concept called the 'autobiography of a smart car'. The idea has been mathematically formalized to concretize its semantic. A threat model is used to derive the system requirements for our proposed system and, finally, we have presented and discussed our system architecture and how it meets all those identified requirements. Overall, if our system is deployed, it will provide a holistic view of data about any particular car and enable car owners to share their data with others in a privacy-friendly way. We believe that it will pave down the way for future research in this domain.

References

Alexopoulos, N., Daubert, J., Mühlhäuser, M. & Habib, S. M. 2017. Beyond the hype: on using blockchains in trust management for authentication. In *Trustcom/BigDataSE/ICESS,* 546–553.

Bartolomeu, P. C. & Ferreira, J. 2019. Blockchain enabled vehicular communications: fad or future? In *The Proceedings of the 88th Vehicular Technology Conference (VTC-Fall),* 1–5.

Cebe, M., Erdin, E., Akkaya, K., Aksu, H. & Uluagac, S. 2018. Block4forensic: an integrated lightweight blockchain framework for forensics applications of connected vehicles. *IEEE Communications Magazine* **56**(10), 50–57.

Cintron, L., Graham, S., Hodson, D. & Mullins, B. 2019. Distributed-ledger based event attestation for intelligent transportation systems. In *International Conference on Cyber Warfare and Security.*

Coppola, R. & Morisio, M. 2016. Connected car: technologies, issues, future trends. *ACM Computing Surveys (CSUR)* **49**(3), 1–36.

Dannen, C. 2017. *Introducing Ethereum and Solidity*. Springer.

Dorri, A., Steger, M., Kanhere, S. S. & Jurdak, R. 2017. Blockchain: a distributed solution to automotive security and privacy. *IEEE Communications Magazine* **55**(12), 119–125.

Dumont, M. What is Blockchain Bloat? https://medium.com/@apollocurrency/what-is-blockchain-bloat-64025f51521a. 4 April, 2019. Accessed on 26 May, 2019.

ENISA. 2016. Cyber security and resilience of smart cars – good practices and recommendations. https://www.enisa.europa.eu/publications/cyber-security-and-resilience-of-smart-cars/at_download/fullReport. Accessed on 25 May, 2019.

Ethereum. https://www.ethereum.org/. Accessed on January 10, 2018.

Ferdous, M. S., Biswas, K., Chowdhury, M. J. M., Chowdhury, N. & Muthukkumarasamy, V. 2019. Integrated platforms for blockchain enablement. In *Advances in Computers (ADCOM): Role of Blockchain Technology in IoT Applications,* **115**, 41–72. Elsevier.

Ferdous, M. S., Chowdhury, M. J. M., Biswas, K. & Chowdhury, N. 2018. Immutable autobiography of smart cars. *Presented at the 3rd Symposium on Distributed Ledger Technology (SDLT).*

Fraga-Lamas, P. & Fernández-Caramés, T. M. 2019. A review on blockchain technologies for an advanced and cyber-resilient automotive industry. *IEEE Access* **7**, 17578–17598.

Greenberg, A. Hackers Remotely Kill A Jeep on the Highway – With Me in It. https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/. 21 July, 2015. Accessed on 25 May, 2019.

Halder, S., Ghosal, A. & Conti, M. 2019. Secure OTA Software Updates in Connected Vehicles: A survey. arXiv preprint arXiv:1904.00685.

Hyperledger Fabric. https://www.hyperledger.org/projects/fabric. Accessed on 20 February, 2019.

Infographic: What Connected Car Data Reveals About You. https://www.pentasecurity.com/blog/connected-car-data-infographic/. 2 February, 2019. Accessed on 25 May, 2019.

Jadhav, A. & Sonpimple, A. Connected Car Market Size, Share & Trends, Industry Forecast, 2025 from Allied Market Research. https://www.alliedmarketresearch.com/connected-car-market. November, 2018. Accessed on 25 May, 2019.

Lei, A., Cao, Y., Bao, S., Li, D., Asuquo, P., Cruickshank, H. & Sun, Z. 2019. A blockchain based certificate revocation scheme for vehicular communication systems. *Future Generation Computer Systems*.

Nakamoto, S. 2008. Bitcoin: A Peer-to-Peer Electronic Cash System. Working Paper.

Paverd, A. J., Martin, A. & Brown, I. 2014. *Modelling and Automatically Analysing Privacy Properties for Honest-But-Curious Adversaries*. Technical report.

Pedrosa, A. R. & Pau, G. 2018. ChargeltUp: on blockchain-based technologies for autonomous vehicles. In *Proceedings of the 1st Workshop on Cryptocurrencies and Blockchains for Distributed Systems*, 87–92.

Roufa, I., Millerb, R., Hossen, M., Taylora, S. O. T., Xua, W., Gruteserb, M., Trappeb, W. & Seskarb, I. 2010. Security and privacy vulnerabilities of in-car wireless networks: a tire pressure monitoring system case study. In *the 19th USENIX Security Symposium*, 11–13.

Sharma, R. & Chakraborty, S. 2018a. "BlockAPP: using blockchain for authentication and privacy preservation in IoV". In *2018 IEEE Globecom Workshops (GC Wkshps)*, 1–6.

Sharma, R. & Chakraborty, S. 2018b. B2VDM: blockchain based vehicular data management. In *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2337–2343.

Sharma, P. K., Kumar, N. & Park, J. H. 2018. Blockchain-based distributed framework for automotive industry in a smart city. *IEEE Transactions on Industrial Informatics*, **15**(7), 4197–4205.

Shostack, A. 2014. *Threat Modeling: Designing for Security*, *61–64*. Wiley. ISBN 978-1118809990.

Shrestha, R., Bajracharya, R. & Nam, S. Y. 2018. Blockchain-based message dissemination in VANET. In *the Proceedings of the 2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS),* 161–166.

Shrestha, R., Bajracharya, R., Shrestha, A. P. & Nam, S. Y. 2019. A new-type of blockchain for secure message exchange in VANET. *Digital Communications and Networks*.

Singh, M. & Kim, S. 2017. Blockchain Based Intelligent Vehicle Data Sharing Framework. arXiv preprint arXiv:1708.09721.

Singh, M. & Kim, S. 2018. Trust Bit: reward-based intelligent vehicle commination using blockchain paper. In *The IEEE 4th World Forum on Internet of Things (WF-IoT),* 62–67.

Smart Car Hacking: A Major Problem For IoT. https://hackernoon.com/smart-car-hacking-a-major-problem-for-iot-a66c14562419. 31 October, 2017. Accessed on 25 May, 2019.

Wang, X., Zeng, P., Patterson, N., Jiang, F. & Doss, R. 2019. An improved authentication scheme for internet of vehicles based on blockchain technologys. *IEEE Access* **7**, 45061–45072.

Woo, S., Jo, H. J. & Lee, D. H. 2015. A practical wireless attack on the connected car and security protocol for in-vehicle CAN. *Proceedings of IEEE Transactions on Intelligent Transportation Systems* **16**(2), 993–10006.