Secure and Tamper-resilient Distributed Ledger for Data Aggregation in Autonomous Vehicles

Sananda Mitra,* Sumanta Bose,† Sourav Sen Gupta,† Anupam Chattopadhyay†
*Department of Computer Science and Engineering, Techno International New Town, India
†School of Computer Science and Engineering, Nanyang Technological University, Singapore
sananda.mitra8@gmail.com, sumanta001@e.ntu.edu.sg, sg.sourav@ntu.edu.sg, anupam@ntu.edu.sg

***Abstract*—The phases those turn the wheels of an autonomous vehicle, are perception, decision and actuation. Among these, the major highlight of recent research has been perception through diverse sensors, and decision through an ever-aggressive cloud/fog/mist computing setup. In this paper, we take a closer look into the flow of data, both internal and external to the autonomous vehicle. We argue that confidentiality, integrity and availability of these data are critical to the eventual adoption of higher-level security and privacy mechanisms in autonomous vehicles. To that effect, we propose a secure and tamper-resilient distributed ledger as an underlying enabler for intra-vehicular data aggregation, and study its security and privacy issues under appropriate adversarial models, where the distributed ledger is instantiated as a standard consortium blockchain.**

***Index Terms*—autonomous vehicles, data aggregation, distributed ledger, blockchain, tamper-resilience, authenticity**

I. Introduction

Autonomous Vehicles (AVs) have the potential to disrupt the landscape of transportation systems through optimized routing, last-mile connectivity, reduced congestion, shared mobility, accident prevention and resource optimization. The industry recognizes six levels of autonomy [1] – Level 0 (no autonomy) to Level 5 (full autonomy), as per SAE taxonomy [2]. We have already reached the prototyping stage for Level 4 (high) autonomy, where all critical operations are autonomous, with provisions for human intervention only in cases of exigency. It is expected that Level 5 (full) autonomy, without the need for any human intervention, will be a reality by 2021 [3].

Autonomy of an AV relies on a synergistic data ecosystem. With the recent proliferation of cloud, fog and mist computing [4], it is most appropriate to consider an AV as a complex cyber-physical system (CPS) embedded within an *intelligent* grid of static (infrastructured) and dynamic (infrastructureless) information agents, where the autonomy of the AV unit may be considered as a three-phase reinforcement learning:

- Perception — Interaction with information agents, within the system and in the environment, for data accumulation;
- Decision — Inference from accumulated data, through artificial intelligence, for (real-time) cognitive resolution;
- Actuation — Implementation of cognitive decisions, with adequate feedback mechanism, for automated operation.

In each of these three phases, *data* and the information derived from it plays a major role in determining the runtime safety, stability and security of the entire AV ecosystem.

Vulnerabilities arising from the context of smart mobility on a fleet of AVs is quite unlike that of conventional CPS. Attacks on an AV, in isolation or within a fleet, may range across data leakage, forged data, identity spoofing, data theft, denial-of-service, and many more as we approach Level 5 autonomy [5], [6], [7]. In fact, the security and privacy issues may affect not just AVs, but also threaten smart mobility grids consisting of roadside sensors, security cameras, traffic signals, toll units, parking facilities, electric charging stations, entertainment systems, smart home networks, diagnostic networks, insurance agencies, service centers and system vendors [8], [9], [10].

In recent works [8], [11], [12], researchers have extensively studied the likelihood and severity of potential threats on AVs by considering the attackers' skill and motivation, vulnerable components in the vehicle, attack surfaces, and eventual repercussions. In contrast, we emphasize the role of *data* and *information* in determining the safety and stability of an AV, and propose a secure data accumulation, transportation and aggregation framework in the context of an AV.

II. Data-Flow in Autonomous Vehicles

In a Level 4 or Level 5 AV, the data flow ranges across a multitude of *information units*, including intra-vehicular control units (●—), perception sensors (●—), and peripheral communication modules (●—), as illustrated in Fig. 1. We model the complexity of the data network in three functional abstractions — data generation (telematics and diagnostics), data acquisition (perception and communication), and data processing (decision and actuation), as explained further.

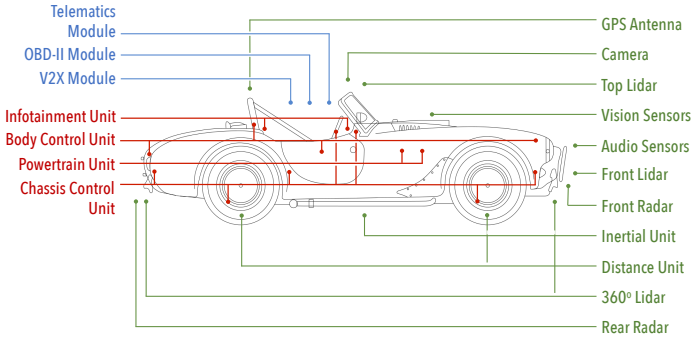

Fig. 1. Information Units in typical AVs: ●— Intra-vehicular Control Units, ●— Perception Sensors, and ●— Peripheral Communication Modules.

A. Data Generation: Telematics and Diagnostics

AVs use various intra-vehicular sensors to improve driving accuracy, and the collective data generated from these sensors are provided to the Electronic Control Units (ECUs) that enable complex driving assistance and in-vehicle comfort [13], [14], [15]. Fig. 1 illustrates the major on-board control units (•—). The *Powertrain Unit* controls the engine and transmission, while the *Chassis Unit* encompasses steering control, anti-lock braking, airbags and adaptive cruise control. The *Body Unit* controls electronic systems related to central locking, lights, air conditioning, while the *Infotainment Unit* takes care of in-vehicle entertainment provisions (radio, media, broadcast) and major information systems like GPS navigation, radio, and telecommunication services [14], [15].

Communication between AV control units is of paramount importance to ensure performance and safety. The standard communication protocols for safety critical (powertrain, chassis) and non-critical operations are CAN, CAN FD, FlexRay, LIN, MOST, and Ethernet [5], [16]. In addition, OBD-II signalling protocols and WiFi or Cellular support exist for cloud, fog and mist access [7], [15]. Smart AV operation requires a number of components inside the vehicle to generate a substantial amount of data, including diagnostics, driving behavior, failure reports, and in-vehicle services [17].

B. Data Acquisition: Perception and Communication

Perception and communication ecosystem of an AV is the primarily driver of vehicle dynamics. Fig. 1 illustrates major perception sensor units (•—) and peripheral communication modules (•—) present in an AV. Communication supported by V2X infrastructure [18] is used to acquire peripheral data imperative for navigation-fidelity. Self-driving cars use a range of perception sensors like LiDAR, RADAR, front and rear cameras, inertial sensors, to name a few [19]. Most sensors lack processing power, and hence the data acquired through these sensors are fused at a heterogeneous data processing framework that can make complex self-driving decisions like adaptive cruise control and automated emergency braking [5], [15]. The fusion of data is also helpful in providing redundancy to cover for the limitations of individual technologies.

Conventional data fusion requires interconnection between internal communication protocols to external communication ecosystem via a common gateway [5], [7]. This poses a potentially serious cyber-security flaw in prevalent AV designs, where the common gateway acts as a single point of failure. Gateway compromise may lead to vulnerabilities in data acquisition and communication within an AV network, resulting in critical failure in safety, security and privacy of the system [5].

C. Data Processing: Decision and Actuation

The AI-driven computing core of an AV is responsible for processing the data generated within the system (Sec. II-A) and the data accumulated from the environment (Sec. II-B) to model the smart mobility grid in and around the vehicle, and take informed decisions regarding control, navigation, safety, maintenance and infotainment within the system [20].

Autonomous Navigation, the primary requirement of an AV, presents quite a challenging scenario for data aggregation and processing with increasing levels of autonomy. While advances in AI address data processing needs, data acquisition and fusion requires a robust cyber-security framework to ensure safety and security [6], [8]. Data flow for autonomous navigation encompasses data acquisition from intra-vehicular control units (powertrain, chassis, body, infotainment) and peripheral communication modules (telematics, V2X) for perception, followed by data fusion and processing at the central AI module for real-time decisions, and finally transmission of control signals to electronic and mechanical control units (powertrain, chassis, body, infotainment, telematics, diagnostics) for actuation [20]. We envisage a distributed ledger as an underlying framework for data aggregation in all such cases of automation to ensure security, privacy and tamper-resilience.

III. Data Aggregation using Distributed Ledger

Secure and tamper-resilient data aggregation within a set of participating entities with a shared state is the need of the hour for AVs, and the forte of distributed ledgers like blockchain. The advantages of using Distributed Ledger Technology (DLT) are many-fold, especially in terms of state-of-the-art data privacy and real-time automation through smart-contracts.

A. Distributed Ledger Technology

Distributed ledgers (such as blockchain) provide a shared state of records or transactions that have been executed through consensus among a consortium of participating entities [21], [22], [23]. The ledger contains a verifiable and tamper-resilient record of every transaction in the network, and provides:

- Provenance of data — verifiable source of origin for all records and transactions, eliminating conflict by design;
- Validity of data — distributed verification of all records and transactions, adhering to some multi-party consensus;
- Immutability of data — secure storage of records or transactions with tamper-proof cryptographic mechanism.

Blockchain, a publicly-verifiable consensus-driven tamper-resilient distributed ledger, forms the backbone of Bitcoin [24] and several other cryptocurrencies. Depending on network structure, identity and trust, blockchain architectures have been classified into three categories, namely public, private and consortium [25]. With the advent of Ethereum [21], blockchain ledgers equipped with smart contracts started modelling a distributed operating system, while Hyperledger [22] and Corda [23] have propelled blockchain into the domain of consensus-driven automation networks. This is precisely where distributed ledgers, especially blockchains, interface with the AV ecosystem in terms of data aggregation and automation.

B. Blockchain for Data Aggregation

We observe that the data aggregation network in an AV resembles a consortium blockchain network, which operates under the federation of multiple groups of entities, with the consensus dependent on a pre-decided subset of participants [25]. In Fig. 2, we present a conceptual model of the

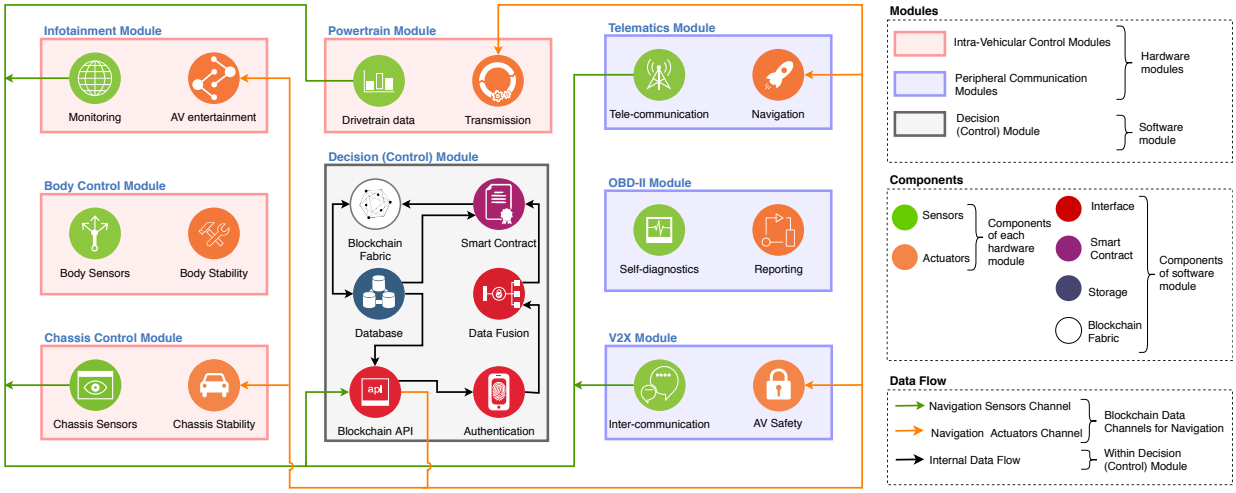

Fig. 2. Components, Modules and Data Flow Channels for *Autonomous Navigation* with an underlying Blockchain framework for Data Aggregation.

Autonomous Navigation data flow in an AV as a blockchain network, with its core components, modules and data flow channels following the operational layout of a consortium [22].

We partition the *Autonomous Navigation* system into two functional units — the *Perception–Decision* unit and the *Decision–Actuation* unit. In our proposed model, the participating modules (powertrain, chassis, telematics, infotainment, V2X) form a consortium structure and contribute components as members to form the data-flow channels. Two data flow channels, *navigation sensor channel* (→) and *navigation actuation channel* (→), control the chain of events during navigation, and record each event in the shared ledger. Each module in the consortium network contributes its on-board sensors and peripheral communication ports towards the navigation sensor channel, and its actuation units and ECUs towards the navigation actuation channel, as shown in Fig. 2.

The channels interact through the central Decision (Control) Module, which is responsible for data fusion and decisions. Raw data from the sensors, authenticated using their individual identities, are recorded in the form of smart contracts for further validation, and the on-board buses (CAN, LIN, MOST, FlexRay, Ethernet) act as mediators to aggregate data from sensors connected to each one of them. This mutually authenticated data, once aggregated, is used by the CPUs to attain a consensus-driven decision. Finally, the decision is communicated to all actuation units and ECUs via the navigation actuation channel, with source verification and recipient attribution using smart contracts to ensure overall security and privacy of the navigation control system.

The consortium blockchain framework, comprising of the distributed database, the authentication module, and data fusion smart contracts, and the two data channels described above, forms the backbone for *Autonomous Navigation*. This framework may be strategically extended to encompass the complete automation landscape of an AV, with multiple data channels connecting the sensors, actuators and communication modules on-board and in the smart mobility ecosystem.

C. Security and Implementation Issues

Security and privacy challenges in an AV ecosystem are multi-fold — protection from data breach, secure decision and control units, integrated embedded security, secure (over-the-air) firmware updates and feature activation, secure V2X communication, secure (remote) diagnosis, privacy-aware data sharing, and many more. Moreover, most of the safety critical operations in an AV demands real-time response, making life even harder for a security designer. In certain scenarios, a distributed ledger based data aggregation resolves a number of existing issues [5], [6], [7], [8], [9], [10], [11], [12] pertaining to security and privacy in an AV, and brings forth the inherent benefits of consensus-driven decentralization.

While the distributed nature of the ledger naturally provides redundancy and elimintates the risk of a single point-of-failure, the consensus amongst the peers in the *navigation sensor channel* and *navigation actuation channel* eliminates the risk of single node vulnerability. This addresses the risks of sensor spoofing [26] and fortifies the single-point attack surfaces in an AV [27], [28]. As an example, consider attacks on the *Autonomous Navigation* system by injection of erroneous data at the GPS end-point [11], [12]. Such an attack will be thwarted during consensus and validation in the *navigation sensor channel* of the consortium, resulting in an efficient fault detection and automated resolution of inconsistency.

The sensors and control units participating in the consortium will require on-board identity management modules to ensure the integrity and authenticity of the data during aggregation. In addition, a trusted execution environment [29], [30] is required at the Decision (Control) Module for secure execution of the core data fusion and automation logic. The authenticated data fusion will provide consistency and accountability guarantees during secure processing of data for training of the on-board AI core. The immutable and validated record of data aggregated across the AV will also act as a reliable source of information for auditors, manufacturers and insurance providers.

D. DLT and Real-Time Operations

Safety-critical AV operations require real-time decisions. On the contrary, authenticated data aggregation using a distributed ledger followed by robust analysis to gain reliable information is time consuming, depending on the latency of the consensus network and the model used for data analytics. In our proposal, we do not mandate the use of aggregated data real-time in any decision feedback loop, thus eliminating the consensus overhead in safety-critical operations. The primary purpose of data aggregation using a distributed ledger is to have a reliable and validated source of information within the AV ecosystem, without affecting its safety-critical real-time operations.

Using the distributed data aggregation framework within a real-time decision loop in an AV will require an extremely high throughput DLT design, supported by a low-latency consensus mechanism like pBFT [31]. One may also consider a modular design of the consortium network (as in Fig. 2) to allow for a highly efficient consensus based on sharding [32].

IV. Conclusion and Future Works

Smart autonomy deserves a smart design. To the best of our knowledge, we propose the first *data* and *information* centric design of an autonomous vehicle architecture, with integrated measures for security, tamper-resilience and privacy. In the long run, one may similarly model the data aggregation framework of a smart mobility grid as an integrated automation network on top of a distributed ledger framework. Such a framework may comprise of roadside infrastructure, electronic road pricing system, parking facilities, electric charging stations, diagnostic networks, insurance agencies, service centers, intra-vehicular network, system manufacturers, etc.

Acknowledgment

This work is partially supported by a fund from ERI@N, the Energy Research Institute, Nanyang Technological University, Singapore. The authors would like to thank the anonymous reviewers of APCCAS 2018 for their kind comments that helped improve the technical and editorial quality of the paper.

References

[1] R. Peterson and D. Glancy, "The Future of Mobility and Shifting Risk," https://www.aig.com/knowledge-and-insights/k-and-i-article-the-future-of-mobility-and-shifting-risk, 2018.

[2] SAE International, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," https://www.sae.org/standards/content/j3016_201806/, 2018.

[3] J. Walker, "The Self-Driving Car Timeline," https://www.techemergence.com/self-driving-car-timeline-themselves-top-11-automakers/, 2018.

[4] A. Corsaro, "Cloudy, Foggy and Misty Internet of Things," in *7th ACM/SPEC on Intl. Conf. on Performance Engg.*, 2016, pp. 261–261.

[5] M. Wolf, A. Weimerskirch, and C. Paar, "Security in automotive bus systems," in *Workshop on Embedded Security in Cars*, 2004.

[6] O. Henniger, L. Apvrille, A. Fuchs, Y. Roudier, A. Ruddle, and B. Weyl, "Security requirements for automotive on-board networks," in *9th Intl. Conf. on Intelligent Transport Systems Telecomm.*, 2009, pp. 641–646.

[7] J. Petit and S. E. Shladover, "Potential Cyberattacks on Automated Vehicles," *IEEE Trans. Intelligent Transportation Systems*, vol. 16, no. 2, pp. 546–556, 2015.

[8] D. Dominic, S. Chhawri, R. M. Eustice, D. Ma, and A. Weimerskirch, "Risk Assessment for Cooperative Automated Driving," in *Proceedings of the 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy, Austria*, 2016, pp. 47–58.

[9] D. Klinedinst and C. King, "On board diagnostics: Risks and vulnerabilities of the connected vehicle," https://resources.sei.cmu.edu/asset_files/WhitePaper/2016_019_001_453877.pdf, 2016.

[10] C. Bloom, J. Tan, J. Ramjohn, and L. Bauer, "Self-driving cars and data collection: Privacy perceptions of networked autonomous vehicles," in *13th Symp. on Usable Privacy and Security, USA*, 2017, pp. 357–375.

[11] S. Parkinson, P. Ward, K. Wilson, and J. Miller, "Cyber threats facing autonomous and connected vehicles: Future challenges," *IEEE Trans. on Intelligent Transportation Systems*, vol. 18, no. 11, p. 2898, 2017.

[12] A. Burg, A. Chattopadhyay, and K.-Y. Lam, "Wireless Communication and Security Issues for Cyber–Physical Systems and the Internet-of-Things," *Proceedings of the IEEE*, vol. 106, no. 1, pp. 38–60, 2018.

[13] W. J. Fleming, "Overview of automotive sensors," *IEEE Sensors*, vol. 1, no. 4, pp. 296–308, 2001.

[14] G.-N. Sung, C.-Y. Juan, and C.-C. Wang, "Bus Guardian Design for automobile networking ECU nodes compliant with FlexRay standards," in *IEEE Intl. Symposium on Consumer Electronics*, 2008, pp. 1–4.

[15] P. H. L. Rettore, B. P. Santos, A. B. Campolina, L. A. Villas, and A. A. F. Loureiro, "Towards intra-vehicular sensor data fusion," in *19th IEEE Intl. Conf. on Intelligent Transportation Systems, Brazil*, 2016, p. 126.

[16] Infineon Technologies, "Your path to robust and reliable in-vehicle networking," https://www.infineon.com/dgdl/Automotive+Networking_2016.pdf, 2015.

[17] Accenture, "Autonomous Vehicles: The Race is On," https://www.accenture.com/us-en/insights/communications-media/autonomous-vehicles-data-challenges, March 2018.

[18] G. Karagiannis, O. Altintas, E. Ekici, G. J. Heijenk, B. Jarupan, K. Lin, and T. R. Weil, "Vehicular Networking: A Survey and Tutorial on Requirements, Architectures, Challenges, Standards and Solutions," *IEEE Comms. Surveys and Tutorials*, vol. 13, no. 4, pp. 584–616, 2011.

[19] G. Péter, S. Zsolt, and A. Szilárd, "Highly Automated Vehicle Systems," *BME MOGI*, 2014.

[20] J. Straub, W. Amer, C. Ames, K. R. Dayananda, A. Jones, G. Miryala, N. Olson, N. Rockenback, F. Slaby, S. Tipparach, S. Fehringer, D. Jedynak, H. Lou, D. Martin, M. Olberding, A. Oltmanns, B. Goenner, J. Lee, and D. Shipman, "An internetworked self-driving car system-of-systems," in *12th System of Systems Engineering Conf., USA*, 2017.

[21] F. Vogelsteller, V. Buterin *et al.*, "Ethereum Whitepaper," https://github.com/ethereum/wiki/wiki/White-Paper, 2017.

[22] The Linux Foundation, "Hyperledger," https://www.hyperledger.org/projects/fabric, 2016.

[23] M. Hearn, "Corda: A distributed ledger," https://docs.corda.net/head/_static/corda-technical-whitepaper.pdf, 2015.

[24] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," https://bitcoin.org/bitcoin.pdf, 2008.

[25] V. Buterin, "On Public and Private Blockchains," Ethereum Blog, 2015.

[26] Y. Shoukry, P. D. Martin, P. Tabuada, and M. B. Srivastava, "Non-invasive Spoofing Attacks for Anti-lock Braking Systems," in *15th International Workshop on Cryptographic Hardware and Embedded Systems, USA*, 2013, pp. 55–72.

[27] S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, S. Savage, K. Koscher, A. Czeskis, F. Roesner, and T. Kohno, "Comprehensive Experimental Analyses of Automotive Attack Surfaces," in *20th USENIX Security Symposium, USA*, 2011.

[28] D. K. Nilsson, U. Larson, F. Picasso, and E. Jonsson, "A First Simulation of Attacks in the Automotive Network Communications Protocol FlexRay," in *International Workshop on Computational Intelligence in Security for Information Systems, Italy*, 2008, pp. 84–91.

[29] Intel Corporation, "Intel Trusted Execution Technology," https://www.intel.com/content/www/us/en/architecture-and-technology/trusted-infrastructure-overview.html.

[30] ARM Limited, "ARM Security Technology – Building a Secure System using TrustZone Technology," https://www.arm.com/products/security-on-arm/trustzone.

[31] M. Castro and B. Liskov, "Practical byzantine fault tolerance and proactive recovery," *ACM Trans. Comput. Syst.*, vol. 20, no. 4, pp. 398–461, 2002.

[32] E. Kokoris-Kogias, P. Jovanovic, L. Gasser, N. Gailly, E. Syta, and B. Ford, "Omniledger: A secure, scale-out, decentralized ledger via sharding," in *2018 IEEE Symposium on Security and Privacy, SP 2018, Proceedings, 21-23 May 2018, San Francisco, California, USA*. IEEE, 2018, pp. 583–598. [Online]. Available: https://doi.org/10.1109/SP.2018.000-5