future internet

MDPI

Article

A Fair Crowd-Sourced Automotive Data Monetization Approach Using Substrate Hybrid Consensus Blockchain

Cyril Naves Samuel [1,2,*], François Verdier [1,*], Severine Glock [2] and Patricia Guitton-Ouhamou [2]

[1] Laboratoire d'Electronique, Antennes et Télécommunications/National Centre for Scientific Research Unité Mixte de Recherche, Electronic Department, Campus Sophia Tech, Université Côte d'Azur, 930 Routes Des Colles, 06410 Nice, France
[2] Renault Group, Technocentre, 1 Avenue du Golf, 78084 Guyancourt, France; severine.glock@renault.com (S.G.); patricia.guitton-ouhamou@renault.com (P.G.-O.)
* Correspondence: cyril.samuel@renault.com (C.N.S.); francois.verdier@univ-cotedazur.fr (F.V.)

Abstract: This work presents a private consortium blockchain-based automotive data monetization architecture implementation using the Substrate blockchain framework. Architecture is decentralized where crowd-sourced data from vehicles are collectively auctioned ensuring data privacy and security. Smart Contracts and OffChain worker interactions built along with the blockchain make it interoperable with external systems to send or receive data. The work is deployed in a Kubernetes cloud platform and evaluated on different parameters like throughput, hybrid consensus algorithms AuRa and BABE, along with GRANDPA performance in terms of forks and scalability for increasing node participants. The hybrid consensus algorithms are studied in depth to understand the difference and performance in the separation of block creation by AuRa and BABE followed by chain finalization through the GRANDPA protocol.

Keywords: blockchain; substrate; hybrid consensus; Polkadot; authority round; BABE; GRANDPA

Citation: Samuel, C.N.; Verdier, F.; Glock, S.; Guitton-Ouhamou, P. A Fair Crowd-Sourced Automotive Data Monetization Approach Using Substrate Hybrid Consensus Blockchain. *Future Internet* **2024**, *16*, 156. https://doi.org/10.3390/fi16050156

Academic Editors: Javier Prieto and Ricardo Alonso

Received: 16 February 2024
Revised: 1 April 2024
Accepted: 17 April 2024
Published: 30 April 2024

1. Introduction

"Data is the new oil" or "Information is the oil of the 21st century, and analytics is the combustion engine" are breakthrough analogical phrase notions adopted by various organizations like Apple, Amazon, Facebook, Alphabet or automotive innovators like Waymo, Tesla, Renault as they embark the digital age of autonomous driving (ADAS). ADAS systems learn and predict using data generated from a fleet of vehicles where each autonomous vehicle can generate up to 300 Terabytes of data per year comprising of different sensors like Radar, Light Detection and Ranging (LIDAR), Camera, Ultrasonic, Vehicle motion, Global Navigation Satellite System (GNSS) and Inertial Measurement Unit (IMU).

These positive connotations require a certain prudence as data need the five Vs of Volume, Velocity, Variety, Veracity and Value and the three As of Analytics, Algorithms and Applications. These have been well argued, as with the explosion of data by vehicles, computing devices, or the Internet of Things, the data need to have practice. These data practices or management needed are

- Data Provenance: Knowledge of the origin of data to generate insights and void bias.
- Data Privacy: Right to access, rectification, erasure, processing and portability guaranteed inside the European Union by the General Data Protection and Regulations (GDPRs) while handling data.
- Data Protection: Securing the data is essential as it comes at a high cost if availability is not maintained as earlier; even the Toy Story Movie Franchise had the risk of becoming obsolete when a technician accidentally deleted it.
- Data Preparation: It is necessary to clean insight extraction to make it AI-usable as data quality need to be augmented.

To ensure the properties mentioned above and the practices, the data need to be incentivized and adopt a decentralized technology like blockchain where multiple stakeholders like data creators, handlers, exploiters and aggregators are involved. In our use case, we can consider the data creators to be vehicles or vehicle proprietors, data handlers to be Original Equipment Manufacturers like Renault, data exploiters like Autonomous Driving Solution Providers like Waymo, Wejo, Momenta or Oxbotica, as well as mobility data aggregators like DIMO a blockchain-based solution, Otonomo, Carscan.

The major contributions of this paper are as follows:

- A data monetization architecture based on bidding and guaranteeing data privacy is proposed, incentivizing the data creator being a vehicle, the intermediary being the vehicle OEM and finally the data consumer being the radar OEM.
- Ensuring the credibility and the validity of the data with a review system enabled through blockchain smart contracts.
- A new blockchain platform, Substrate, is chosen for implementing the solution which has the features of embedded Smart Contracts, hybrid consensus of block proposal and finalization algorithms.
- The solution is evaluated for its performance on a cloud platform to measure its throughput, consensus forks, and finality of the transactions.

The paper is organized as follows:

- The first part introduces data monetization and its importance as well as different commercial decentralized mobility solutions.
- Next, we follow with the state of the art on data monetization and evaluation as well as comparison for each of the works. It is followed by a discussion of different mobility data standards.
- We then continue with our data monetization architecture, blockchain platform, and hybrid consensus. We continue with the data monetization use-case definition, architecture proposal and implementation details on the cloud infrastructure.
- We then end with the functional, performance evaluation with more focus on the hybrid consensus algorithms and a conclusion with future work proposals.

1.1. Data Monetization

Mckinsey analyze incentivization through data monetization, which states it is a differentiator and is still nascent. They also assert that with the available data, it is necessary to engage with other partners to create new business ecosystems, and it is essential to dissolve sectoral borders. This is one of the business models we would approach in our architecture to create an ecosystem of related businesses. Data monetization defined by Gartner as "using data for quantifiable economic benefit" can be one of the following strategies as highlighted in [1] along with the adopted organizations:

- Asset Sale: Sale of Direct Data by Strava, Verizon Wireless.
- Business Process Improvement: Value is extracted from data for optimization of one's business process by Lufthansa, ThyssenKrupp, and Deutsche Bank.
- Product/Service Innovation: Offering new business services or processes based on data by IBM, Rolls Royce.
- Product/Service Optimization: Optimization of existing service based on data by Ford, Zara and Pirelli.
- Data Insight Sale: Selling derived knowledge by analytics, visualization by Olery, Sendify and DealAngel.
- Contextualization: Addition of supplementary over the existing data for economic benefits by Staples and Walmart.
- Individualization: Customer data are used to customize the product offering and preferences, enhancing the value proposition by eBay, Daimler or Netflix.

In our use case, we adopt data bartering indirectly as the strategy process between the different ecosystem actors. It is a simplified approach, not considering the capabilities

of Big Data or Artificial Intelligence Services, which can be very well included in future scenarios for enhanced returns. We concentrate on an extensible, generic architecture that can be recast and shaped vertically, pipelining other technological services or products. Some challenges in data monetization are final data usage policy, legal liability, cross-border data trade, security, and privacy challenges as uncovered in [2], which we also analyze in our architecture.

1.2. Decentralized Mobility Solutions

In this section, we compare and contrast two decentralized mobility data solutions already in production with a modest market share thriving in an ecosystem of partners for offering "Return on Data" shared with these networks.

- Digital Infrastructure for Moving Objects (DIMO): [3] It is a data-driven decentralized public IoT platform that requires the vehicle owners to install an AutoPi telematics unit in their vehicle, which then submits the data at frequent intervals to the cloud infrastructure. Its infrastructure comprises the Polygon Blockchain, the Inter-Planetary File Storage (IPFS) decentralized storage, the Helium decentralized LoRaWAN wireless network and other additional protocols. Users can submit their data and can gain DIMO tokens as rewards. The platform then utilizes the collected data for receiving recommendations on preventive vehicle maintenance, service history, and other sensor records. The shared data are then sold to aggregators who reward and incentivize the DIMO network and vehicle owners. The compromise we notice with this system is that the data life-cycle is questionable regarding its provenance, destruction, obfuscation by removing sensitive details and the overhead of installing a device for protocol communication. Though the vehicle Users and OEMs monetise their raw data directly they lose the opportunity by failing to enhance the data and build services on top of it which might increase their revenue. Although it is designed to solve the problems of data centralization and privacy concerns, it needs more analysis, as the data are stored by centralized actors in the ecosystem.
- Ocean Protocol: The Ocean Protocol [4] is a data marketplace built on the parity Ethereum Proof of Authority blockchain protocol. It is a decentralized data exchange where individuals or enterprises exchange the data shared as ERC721 data NFT tokens by Ocean smart contracts. Then, ERC20 data tokens are generated for the data service to access the published data for a dynamic or fixed price. The existing use cases generated are for connected living, where the elderly patient's data are shared for customized insurance and medical assistance. Also, the health data for heart condition patients are shared with Roche Diagnostics from a CoaguChek IN Range device to the ocean protocol, enabling data discovery for other third-party partners offering medical services. Also, a data NFT can be marked as purgatory or unusable if there is an issue with data privacy, quality or copyright infringement. To preserve the confidentiality of the data, the marketplace offers compute-to-data, where instead of the data transfer, the buyer of tokens can obtain the pre-computed trained data model to be used, and the raw data stay in the storage of the data marketplace. This protocol is quite comprehensive with no additional hardware requirements, but there is a problem with the data certification, as data shared can be tracked for their provenance. Still, data quality and genuineness are pointers to be considered by them as anyone can share data and earn tokens without pre-emptive checks.

Our data-monetization architecture solves the concerns mentioned above regarding data provenance and certification using a streamlined data flow with no additional hardware required. It also ensures the privacy of the data within a consortium blockchain of agreed and verified partners with mutual benefit for everyone in the network.

1.3. State of the Art

In this sub-section, we explore the literature around data trading services using blockchain, which discusses data trading, and sharing data securely, respecting the user's privacy concerns.

A Decentralized Review System for Data Marketplaces [5] hosting urban quality of life data, vehicle data and other IoT data is designed to replace the conventional centralized rating systems. The Decentralized Data Marketplace consists of the critical elements wherein the buyers and sellers of data interface through the marketplace for data transfer, buying, querying the data, reviewing the data, and providing ratings to the data. The system allocates pre-determined credible reviewers against a set of products in a randomized and double-blinded method to minimize actor collusion and gaining illegal advantage out of any transaction. Reviewers are incentivized through game theoretical modeling, where they identify the conditions for Nash Equilibrium policy for the reviewers. The Nash Equilibrium policy aims at finding an optimal solution to a problem with competitive participants. The system is presented as a general framework and analyzed theoretically with no decentralized platform implementation. The presented work is simulated using NashPy Library, where as long as sufficient incentives increase the probability of a full review, the review quality is enhanced. But there are still some pertinent questions unanswered in this work: (1) the formation of a review committee when the platform is still new and the absence of reviewer credentials, (2) the strategy to review a product continuously or at a determined time frequency, (3) the issue of preservation of the seller's privacy behind the transaction process and (4) the scalability potential of the system with increasing products, reviewers and sellers.

A credible data trading system for IoT data minimizing the risk of fraud and transactions is proposed in [6]. The system permits the data producers and consumers to agree on data and settle the payment on the chain. A credit mechanism is developed to lower the fees incurred during the private Ethereum network implementation participation process. This system's objectives are Consumer Fairness, where customers do not pay for data not received; Producer Fairness, ensuring the minimal risk of data loss; and Privacy, limiting data visibility at a minimal operational cost. The system is evaluated regarding gas consumption and incurred blockchain transactions for the trading scheme, as the idea is to minimize the transaction cost related to transaction gas complexity. Around 35,000 units of gas is estimated for fund deposit transactions and higher for receipt transactions as it needs data signature checking.

A consensus-based distributed auction scheme for data sharing is proposed in [7] for privacy preservation and avoiding malicious collusion of actors. The scheme allows for the participants to group into clusters for privacy and then reach a consensus. The mechanism is further incentivized to share data without privacy leakages. Differential privacy, symmetric encryption and zero-knowledge proofs are incorporated to design the auction mechanism for a trade-off between privacy preservation and social efficiency. The consensus algorithm is constructed where different kinds of witnesses are selected using anonymous verifiable random functions. It is performed without peer interactions and different parallel operations by varying witnesses to verify the proof and result, ensuring finalization. The evaluation of the system shows that the participants reach a consensus on the auction result with low computation and communication costs.

A novel proposition of data marketplace design that satisfies all desirable properties in any system of fairness, efficiency, security, privacy, and regulation adherence is proposed in [8]. The authors implement FairSwap [9] to guarantee fairness where the participants agree on a value wherein a Proof of Misbehavior is generated to punish in case of wrongdoing. The authors rely on any generic encryption and hash function, including Zero-Knowledge Proofs, for transparency, security and privacy. They mention using codified language, such as smart contracts, for the regulation aspect. To maintain efficiency, they suggest the usage of Zero-Knowledge Succinct Non-Interactive Argument of Knowledge (SNARK) or Merkle tree over boolean circuits to analyze encrypted data. The work is a

more theoretical discussion rather than a concrete implementation with particular work on its challenges. The uncovered inconveniences are global fairness issues where fairness in a larger context cannot be guaranteed if more participants of varying interests participate in the ecosystem. Another issue is the inefficiency of predicate checking in encrypted data for arbitrary logic. The incentivization mechanism is not theoretically evaluated, and capturing practical constraints in the framework is complicated. Also, data duplication is quite a problem, and the authors suggest using convergent encryption without revealing any data information.

A proof-of-concept work for creating a novel secure and fair reseller market is proposed in work [10]. It creates a level of trust between trading parties and it avoids the possibility of counterfeit goods. The authors use non-fungible tokens for authenticating the items as well as product ownership leveraging Web3.

1.4. Decentralised Mobility Data Standards

This section explores the standards around the blockchain data market by different working groups of consortiums and private and technical organizations.

1.5. Mobility Open Blockchain Initiative (MOBI)

It is a global smart mobility consortium [11] of mobility providers, technology companies, governments, and NGOs. The consortium has working groups for Vehicle Identity, Usage-Based Insurance, Electric Vehicle Grid Integration, Supply Chain, Finance, Securitization, and Smart Contracts. The authors of the cited work define the Connected Mobility Data Marketplace (CMDM) standards for Vehicle-to-X (everything including cloud, infrastructure, or vehicle). This enables vehicle or fleet owners to monetize the data by selling to third-party providers like road conditions or algorithm developers, ensuring privacy and security. The authors aim to create a standard where infrastructure like roads, bridges, vehicles, and other third-party intelligent transport providers interconnect seamlessly. As data sets by an individual organization are incompatible with one another, this standard creates the interoperability for scalable data sharing ensuring efficient monetization.

1.6. European Telecommunications Standards Institute (ETSI)

As mentioned in Ref. [12], it is a non-profit standardization organization that has published a specification for a permissioned distributed ledger in supporting distributed data management. It defines the specification for data discovery, collection, storage, sharing and computation. It ensures the following requirements: decentralization, trust, incentivization, data provenance, data privacy, integrity, control, sovereignty and data management automation, ensuring GDPR compliance.

1.6.1. International Standard Organisation/TC 307

This Standard TC 307 [13] is used for blockchain and electronic distributed ledger systems and application, interoperability and data exchange between users. It prescribes the data flow model for DLT use cases and decentralized identity standards. It also enlists smart contract security good practices and a data interoperability framework.

1.6.2. European Union Blockchain Observatory and Forum

In report [14], the relationship between blockchain and the automotive sector is discussed for communication and data transaction security. The report is about the supply chain management for components and the introduction of new services that can be built upon. The authors discuss the necessity of data privacy through zero-knowledge proof, privacy-preserving transaction settlement and data interoperability.

2. Significance of Data Monetization Architecture

In line with the above concerns of standards and previous data monetization solutions, we analyze here the necessity for building a new architecture, especially in the mobility context, as follows:

- Vehicle OEMs are hit by the wave of Industry 4.0 with the advent of Digital Twin, where all the configuration and data of a vehicle are stored in the cloud. In close accordance is the launch of a Software-Defined Vehicle (SDV) technology. [15] allowing to "centralize data with other components of ADAS, bodywork, chassis, and telematics in a Physical computer unit". Also, all the customization and upgrades are made over the FOTA (Firmware Over The Air) technology, which is simplified by at most two High-Performance Computers (HPC) for SDV in a vehicle. All these Vehicle-to-Cloud communications, as well as the data that are shared, create the necessity of moving towards a decentralized data monetization where each participant in the ecosystem of OEM like Renault, Component Manufacturers like Qualcomm, Cloud Service providers like Google and the vehicle owner can each own a piece of the pie (service built on top of data and rewards) offered by the data, ensuring mutual benefit.
- Data which the vehicle owner engender have to create a virtuous cycle where each piece of data, starting from the sensor data, can be used by the equipment OEMs like Bosch, Continental. for improvisation, Vehicle OEM can enable the provisioning of data and a vehicle owner can obtain service benefits and up-gradation from the equipment OEM. All these actors create a virtuous cycle of fidelity and continuous improvement complemented by data monetization.
- We design our architecture from a consortium point of view with benefits shared with each actor and guarding data privacy, provenance, certification, accountability and validity with the closed set of participants, which is needed for an automotive ecosystem.

2.1. Next-Generation Distributed Ledger

Our architecture implementation consideration aims to solve the problems encountered in our previous works [16–18] as well as previous discussions as follows:

- Consensus Scalability: As we noticed in our previous works related to Byzantine Fault Tolerance (BFT) blockchains except Clique and QBFT, other protocols' performance was affected as nodes were increased even from a consortium setting of limited participants. Also, in Clique, the finalization of transactions was a questionable aspect, and from an automotive context, we need finalized transactions as security cannot be compromised.
- Interoperability: Blockchain solutions we designed earlier were not interoperable with other blockchains limiting their communication to receive external information and create new synergies and services, which is needed from the current Web3 view standpoint.
- Embedded Smart Contract: Ethereum smart contract, which we used earlier to create ERC20 and ERC721 tokens, used an EVM (Ethereum Virtual Machine) platform. It uses EVM for the interpretation, execution and finalization of smart contract transactions, which introduces additional processing bottlenecks apart from the transaction consensus and processing. But we envisage building the smart contract and the blockchain node binary to avoid separate execution in an EVM for faster and more secure processing.
- Secured Oracle and OffChain Communication: Ethereum-based smart contracts need a third-party decentralized Ethereum-based Oracle for any external communication or knowledge feed to the smart contract decision logic. These are Chainlink, Band protocol, Pyth Network, and others, which are third-party networks to be depended upon, which defeats the purpose of a decentralized solution. So we need an OffChain worker solution that can connect seamlessly to our internal and external Application Programming Interfaces (APIs) or blockchain networks in the form of Oracles, which we solve in our architecture.

- Mutual and Divisible Monetization: Data that need to be monetized have to be beneficial and sustainable by attributing rewards for everyone in the ecosystem and creating new services and enhanced customer experience, value addition, and product evolution.
- Cycle of Data Certification and Provenance: Data that are shared have to be collated from multiple actors or devices, processed, cleaned, obfuscated for privacy concerns and finally submitted via an API or data pool even for a simple raw data monetization without any computation. These successive stages must be acknowledged from their nascent stage until the end stage by certifying at each step as a Signature or Hash and submitting to the decentralized protocol, ensuring a data certification and provenance cycle.
- Privacy By Design:1 Our architecture is designed to follow Article 25 of GDPR principles of "Privacy By Design" rather than "Privacy By Default" [19,20]. In our earlier work on data certification, we followed Privacy by Default through pseudonymization. In this work, we ensure that during data monetization, we follow Privacy by Design, where the data can be read, accessed and verified only by the necessary participants without being public to anyone in the ecosystem.

2.2. Blockchain Framework by Design: Substrate

In our earlier works, we made the architectural decision of Ethereum Bbockchain, a second-generation blockchain comprised of smart contracts to build a distributed application. Meanwhile, Bitcoin is considered the first generation of blockchain as it is focused on the crypto-money transaction with no additional mechanism to create an application or service [21]. But for this design, we choose Substrate [22], which is a third-generation blockchain like Cosmos, Polkadot, Cardano or Avalanche. But to be precise, it is not a precompiled blockchain node but a "modular, extensible framework" that we can utilize to create our custom blockchain node and generate its binary. It can be made interoperable with other chains and also scalable by using the feature of para-chains. Further smart contracts are directly built onto the blockchain node instead of external deployment on EVM. It is built using Rust language and has a core library component called Framework for Runtime Aggregation of Modularized Entities (FRAMEs), which can be used to inherit or build custom blockchain components on top of it. Also, there is a separation of responsibility design by distinguishing node usual activities of cryptography, network or consensus and custom logics of business conditions, external communications like Oracles and other interfaces. All the signed transactions executed in the smart contract pallets can be invariably called Extrinsic in Substrate terminology. Many enterprise networks like Chainx utilize the Substrate framework based on BABE and GRANDPA consensus, Aleph based on a custom BFT algorithm, and Acala, Plasm, Edgware, Moonbeam and Darwinia based on either AuRa or Nominated Proof of Stake customized with GRANDPA consensus algorithms on public networks.

2.3. Hybrid Consensus

In this section, we explain the novel concept of hybrid consensus proposed in Substrate, which comprises two consensus phases as follows:

- Block Authoring: It processes the transactions or extrinsic values and packages them into blocks by the set of validators for each discrete time slot. This is usually performed by a chosen validator in a round-robin selection by Authority Round (AuRa) consensus or Blind Assignment for a Blockchain Extension (BABE) scheme. These blocks are then subjected to the addition as agreed chain storage where a block header contains a reference to its parent or a previous block.
- Block Finalization: A previously authored blockchain contains a chain of blocks and can be subjected to forks where two blocks refer to a single parent block. So to resolve the fork or finalize the chain, we need an additional algorithm to select the best chain, which is the longest chain with higher weightage as in Greediest Heaviest Observed Sub-Tree-based Recursive Ancestor-Deriving Prefix Agreement (GRANDPA). This

ensures a deterministic finality where a chain can never be compromised, as only block authoring offers a probabilistic finality.

Cross-Consensus Messaging (XCMP)

In addition, the Substrate framework is also inter-operable, with a mechanism for communicating between chains called Cross-Consensus Message Passing (XCMP), where any custom message can be exchanged between the chains or parachains as called in Substrate terminology. They are classified into asynchronous, which is characterized by a request-response without any time guarantees, absolute, where the message is exchanged in order and efficiently, asymmetric, where there is no response back to the sender, and agnostic, which has no assumption about the message passed. In our architecture, we do not include these messaging systems as we construct a single unified chain focusing on consensus performance with the capability of extending between chains or parachains in the next version.

2.4. Authority Round (AuRa)

This consensus algorithm [23] has a list of authorities or validators where the block authorship happens at a time slot or step occurring in every time interval. For each step s, a primary node is assigned in a round-robin methodology which proposes a block as highlighted in Figure 1. The node selection is based on a modulus operation of $SmodN$ where each step is S and N is the number of nodes. The proposed block is then accepted by the remaining validators, where only a single selected validator can exist per slot. In case of time drift or network synchronicity, the node time slots can overlap, which can cause multiple proposers per unique time slot. It can cause forks in the chain that need to be resolved by an additional finalization mechanism offering only probabilistic finality. The communication complexity for this protocol is $O(n^2)$ when considering the block acceptance phase.

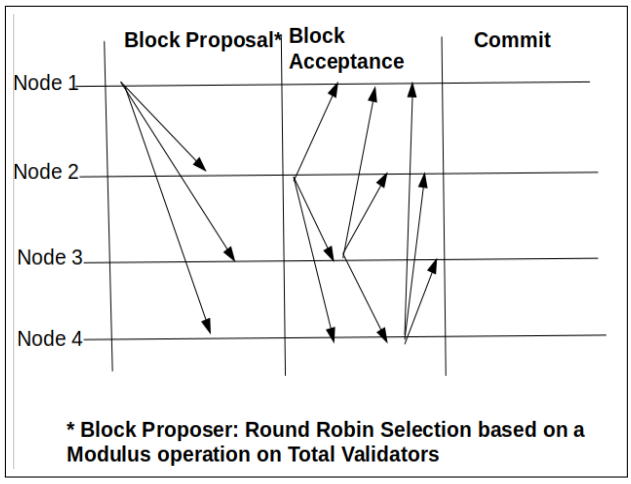

Figure 1. Authority Round (AuRa) Consensus.

2.5. Blind Assignment for Blockchain Extension (BABE)

Block production mechanism BABE [24] is inspired by Ouroboros-Praos [25], a proof-of-stake consensus algorithm. It is also a time slot-based algorithm, as highlighted in Figure 2, organized as epochs. Each epoch comprises a set of successive time slots at pre-configured intervals. It overcomes the security threat in the AuRa protocol, where a validator at a particular block height can be predicted and can be subjected to attacks as it is based on a modulus operation. The selection of the primary validator is based on a verifiable random function (VRF) with the input of a commonly agreed randomized seed, the current slot number, and the validator's private key. If the VRF output generated by a validator is below a commonly agreed threshold δ, then it is chosen to be the primary

validator for the slot. In parallel, a secondary validator is selected as a fallback if the primary one does not respond for its eligible time slot. The other validator nodes then agree upon this proposed block. But it has only a probabilistic finality and needs to be added with the GRANDPA finality gadget as the algorithm is termed to transcend as a deterministic finality. The communication complexity is O(n) as there is a block proposal phase to broadcast the block.

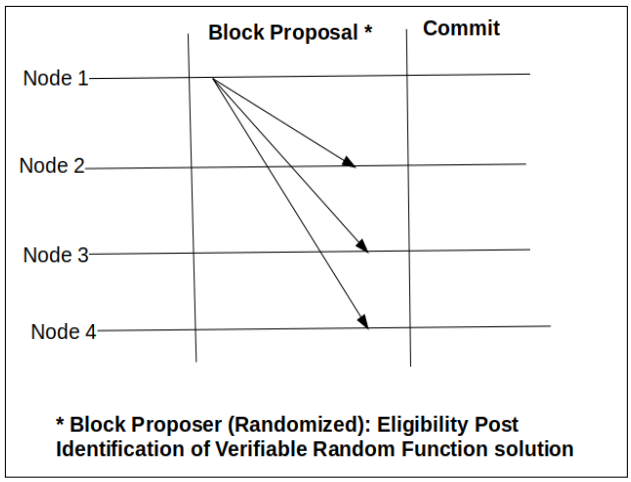

Figure 2. Blind Assignment for Blockchain Extension (BABE) Consensus.

2.6. Greediest Heaviest Observed Sub-Tree-Based Recursive Ancestor-Deriving Prefix Agreement (GRANDPA)

The GRANDPA finality gadget [26] chain finalization protocol can be coupled with AuRa or BABE consensus previously explained for rendering the chain a deterministic finality in cases of forks due to network partition or malicious behavior. As represented in Figure 3, it is a protocol where the validator agrees on the chain and not the blocks. The authors apply votes transitively until the block number with the highest votes is chosen to be final. This, in turn, allows for the chain of blocks, i.e., a chain, to be finalized at once in a single round. In this protocol, a validator is selected to be the primary broadcast of the highest block. Then, during the "pre-vote" phase, each validator endorses a particular block height, confirmed upon a supermajority (2/3) of validators. Then, based on the "pre-vote" previous round, each validator casts a "pre-commit" vote upon which they finalize the chain. The communication complexity is $O(n^2)$ based on the messages exchanged in the pre-vote or pre-commit phases.

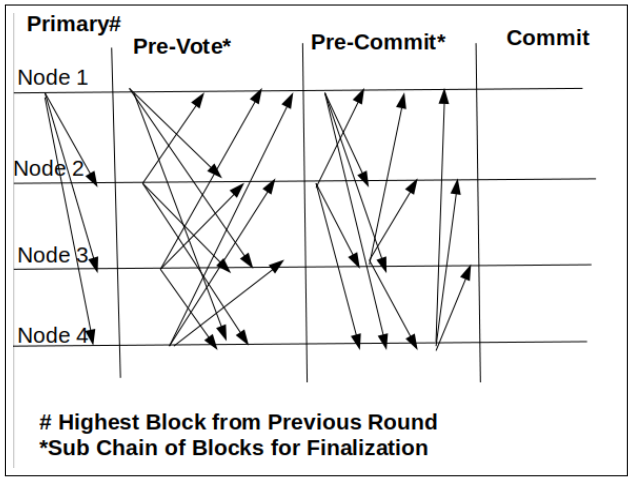

Figure 3. Greediest Heaviest Observed Sub-Tree-based Recursive Ancestor-Deriving Prefix Agreement (GRANDPA) Consensus.

The GRANDPA protocol working mechaism is explained in Figure 4, where the encircled black block represents the recently finalized chain. While considering the four forks signified by yellow sub-chains, each block has either 2 or 1 as weightage, with fork described as Fork 1, Fork 2, Fork 3, and Fork 4. The primary block is signified with Weightage 1, and the secondary one is signified as Weightage 2, which indicates either the block proposition by a normal proposer or a fallback proposer. The GRANDPA algorithm chooses Fork 2 as it has the longest chain with the highest primaries and is built on the last finalized block. So the factors of chain length, the previous finalized block, and the most primaries are chosen despite Fork 4 having the longest chain with the most primaries but not built on the finalized block.

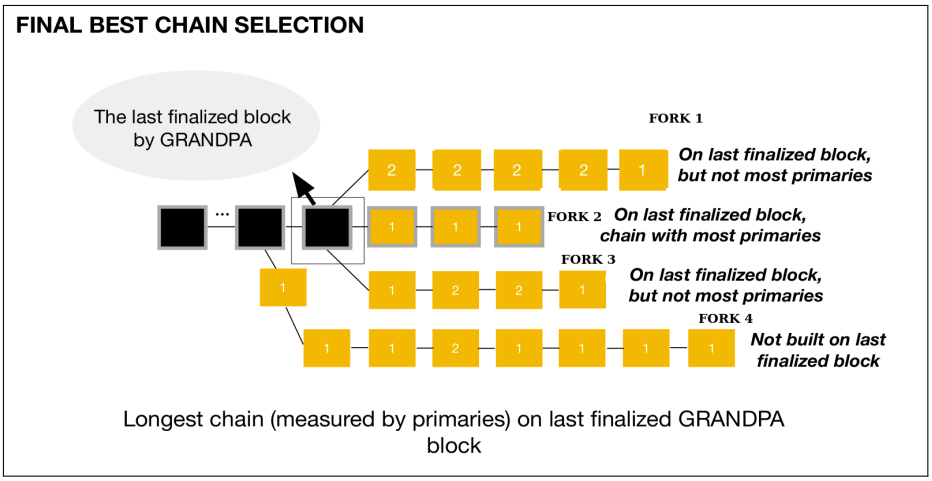

Figure 4. GRANDPA Consensus Finalized Chain [26].

3. Use-Case Definition

In this section, we elucidate the use case of creating an automotive radar data value chain by incentivization and the fidelity mechanism. RADAR (Radio Detection and Ranging) data are collected and consumed based on crowdsourcing from numerous vehicles. In the reverse sense, it is offered as a radar data service by enhancing and building map road radar signature data service using it.

3.1. RADAR Significance

Automotive RADAR furnishes the essential range and speed data for driver assistance systems, including Long-Range RADAR (LRR) adaptive cruise control, automatic emergency braking, cross-traffic alert, lane change assist, and Short-Range RADAR (SRR) parking aid, as well as pedestrian detection. The signal processing functions involved are range estimation, Doppler frequency estimation, direction of arrival estimation, and tracking. In normal cases, multiple RADAR channels are required for angular information or to compensate for any information loss. In our case, we adopt ocalization-based data collection from multiple vehicles for data fusion and extract any valuable information. The significance of the automotive RADAR market is increasing at 40% year-on-year, especially in the premium or mid-segment vehicles. Prominent Automotive RADAR Original Equipment manufacturers include Bosch [27], Continental, NXP Semiconductors or STMicroelectronics for object detection even at 200 km/h. Also, for vehicle safety concerns, as we evolve from L2 autonomous vehicles to fully autonomous L5 vehicles, RADAR is necessary for safety optimization while driving.

3.2. Road RADAR Signature

Road signature [27] is a crowd-sourced localization service for autonomous vehicles to detect the relative position of other vehicles and objects in their environment for accurate and reliable localization. As represented in Figure 5, the road signature is crowd-sourced

or vehicle-sourced in the vehicle surrounding comprising lane markings, gas stations and guard rails within a decimeter range of the vehicle. The vehicle generates RADAR and video sensor data while in motion, which are collated at determined intervals via the Telematics Control Unit of the vehicle to the cloud of RADAR OEM. RADAR OEM observes all these received data of RADAR and video to extract the object details and regenerate the environment based on the data. Video and RADAR sensor data complement the rendering of a localization environment. As discussed earlier, the RADAR OEM, e.g., Continental or Bosch, integrate these data with map data to produce an updated and globally consistent map. The generated road RADAR signature high-definition resolution map, which is offered as a subscription service back to the vehicle user community, comprises three layers:

- Localization layer: Vehicle position is determined based on its driving lane and merged with RADAR road signature, including the video localization map. Also, the object information in the vicinity is compared with the information from the other sensors to determine relative position.
- Planning Layer: This layer provides driving planning information comprising the road course, traffic sign and speed limit, including bends and gradients.
- Dynamic Layer: Traffic information, including deadlocks, construction work, maintenance or parking space availability, is provided.

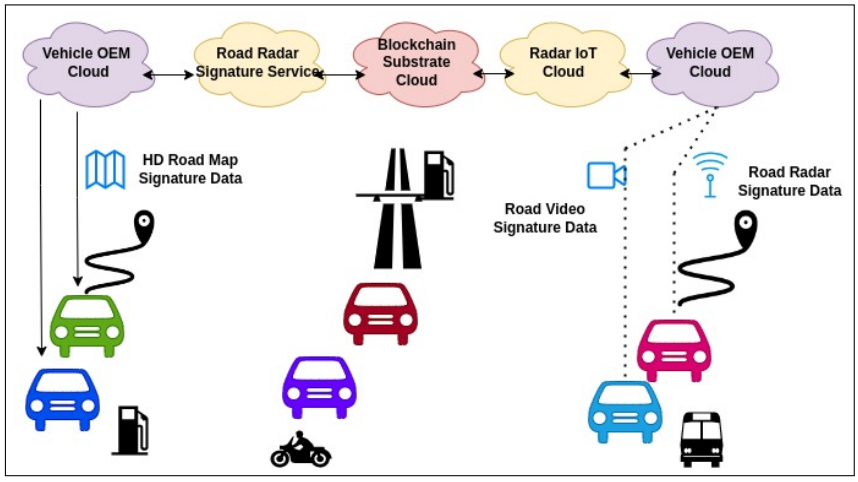

Figure 5. Automotive Road RADAR Signature Use Case.

3.3. Virtuous Cycle of Road RADAR Signature Data

As represented in Figure 6, the road RADAR signature workflow is facilitated and augmented with acknowledgment, data certification, transparency, privacy, dynamic pricing mechanism accompanied by fidelity and incentivization mechanism. Our data monetization use case involves the following components realized through the development of pallet smart contracts on the Substrate blockchain framework:

- Asset Component: This component creates an asset for vehicle OEM as a non-fungible ERC 721 token in the network. RADAR data of the vehicles for a defined condition of localization, time period, error tolerance and other miscellaneous metadata are represented as an asset along with its floor and ceiling price.
- Asset Service Component: The component is similar to the asset component but consecrated towards the road RADAR signature map, which is offered by the RADAR OEM to the vehicle users for enhanced safety, autonomicity, and optimized localization information which is accompanied by a ceiling and floor price.
- Bidding Component: Each asset or asset service created as an ERC 721 token has its price determined based on a transparent bidding process between RADAR OEM, Vehicle OEM and Vehicle Users consenting to the transaction.
- OffChain Component: As soon as the asset or the asset service is sold, data must be offloaded from each vehicle to RADAR OEM via Vehicle OEM. Vehicle OEM acts

as a guarantor or mediator in the blockchain for initial data collation, data cleaning and obfuscating of any sensitive or private data. These data processing steps are performed offchain in a cloud middleware but closely looped with the blockchain to certify the process and the data.

- Asset Data Collation and Certification: The data offloaded to the RADAR OEM are then processed, analyzed and interpolated with map data to be offered as a service later completing the asset finalization along with certification along its entire process.
- Asset Service Offering: Road RADAR map signature is offered to the consenting vehicle back in the blockchain ecosystem with an interest of subscriptions collated in the ledger by Vehicle OEM and transaction completion post the fund transfer.
- Incentivization for Everyone: For each transaction of asset or asset service involving RADAR data monetization and Road RADAR Signature Service, each actor in the ecosystem is attributed an incentivization. The asset monetary transaction distributes the commission to Vehicle OEM, Vehicle Users, and RADAR OEM. The necessary monetary benefit along with data certification, anonymity and quality are maintained. Moreover, the RADAR data or RADAR signature service represented as tokens ascend in price appreciation based on usage review submitted to the smart contract.
- Cyclical Economy Monetization: In the proposed architecture, the vehicle owners can monetize their data generated in the form of credits which can be later reclaimed or used in subscribing to any services like insurance and maintenance offered by Vehicle OEM. Also, the advantage for Vehicle OEMs is creation of repeated sales of products and services benefitting the customer as well as the organization subsidizing the cost of building a vehicle connectivity infrastructure. RADAR OEM can improvise their services and products leveraging from the vehicle owners data. The data can be used to train and optimize the services offering an intelligent and custom-tailored solution to each vehicle in the form of traffic advice, road conditions, driving behavior, etc.

This completes the explanation of the virtuous data cycle since its inception as an asset from the vehicle to its collation in the RADAR cloud. Then this asset is utilised to create an asset service for the vehicle users based on the shared data imbibing fidelity and sustained benefit for everyone in the monetization ecosystem.

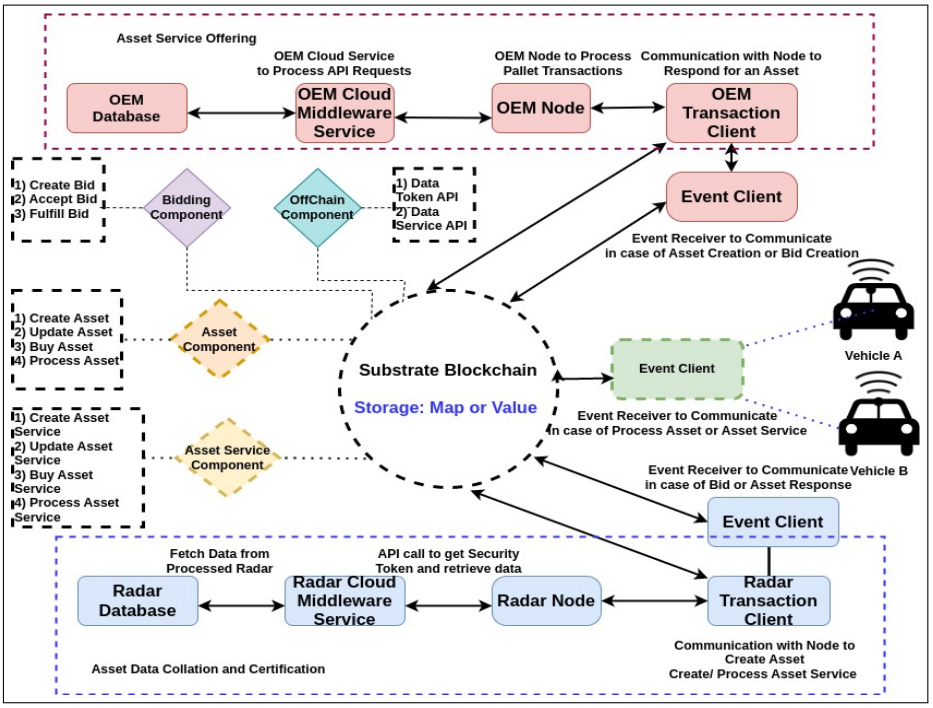

Figure 6. Automotive Data Monetization Architecture.

4. Architecture Solution

This section examines the architecture flow in terms of a smart contract transaction sequence diagram for the non-fungible token-based asset and asset service proposition scenarios. The blockchain ecosystem network comprises nodes representing Vehicle_OEM, and RADAR_OEM as validators with the possibility of extending the membership to other members interested in data; for example, the map provider or the government service as it deals with traffic, road and toll maintenance. The constructed blockchain network is of a private consortium type, where each participant is aware of the public key of the other actors, as each one has a crypto wallet secured by a public cryptography system. In all these explanations, we consider the following principal actors:

- DataMarket: Decentralized Smart Contract Pallet realized on a Substrate blockchain framework, which orchestrates the entire data monetization transaction subject to hybrid consensus algorithms and its validation.
- Vehicle_OEM_X | Y: Vehicle Original Equipment Manufacturers like Renault, Stellantis or Volkswagen, which provide the OffChain cloud infrastructure and participate as validators in the blockchain consensus. Each participant installs and maintains the blockchain client for data transfer, transaction execution, and liaison of the vehicles.
- Vehicle_M | N: Vehicle Users or Autonomous Fleet devices have a wallet-based account in the blockchain process and provide data necessary for monetization and subscribe to the service offered for enhancing driving.
- RADAR_OEM_A | B: RADAR Original Equipment manufacturers like Continental, Bosch and NXP are interested in the generated RADAR data from the vehicles for enhancing their product offering as well as creating additional business services for customers like road RADAR signature map for their end users.

4.1. Tokenized Non-Fungible Asset Data Set Component

Smart Contract transactions involving the OffChain Sequence for the NFT asset component involving RADAR data transfer are represented in Figure 7 for the initial bidding process, and the asset finalization process is explained as follows:

- Data Demand Phase: In the initial phase, when the RADAR_OEM needs the RADAR data, it publishes the demand as an event notification transaction onto the network with the location, vehicle speed and acceptable price limit requirements. Vehicle_OEM can then respond to the event by broadcasting the demand to its vehicle clients for its participation contentment.
- Data Asset Offering and Bid Phase: Vehicle_OEM_X | Y responds individually by creating an ERC 721 token as an asset with asset criteria, floor, and roof price. Then, to maintain bid privacy, RADAR_OEM submits an encrypted bid with the public key of Vehicle_OEM_X | Y, respectively. Then, among the bids, one of RADAR_OEM_As is accepted by Vehicle_OEM_X as the bid of RADAR_OEM_B is not an acceptable price. The fund of RADAR_OEM_A for the bid is transferred to the escrow account as an intermediate transfer which is transferred to Vehicle_OEM_X when the bid criteria and the asset are respected along with necessary certification proof submissions.
- Data Aggregation and Certification: Vehicle_OEM_X starts the data collection job from its set of vehicle clients with the necessary criteria. Vehicles start recording RADAR data and submitting it with its signature hash as proof to the data pool and smart contract. Then, Vehicle_OEM_X processes the data removing the user's sensitive identification data, submits the final hash proof to the ledger, and generates a data access token to a middleware service to be retrieved later.
- OffChain Data Transfer and OAuth Access: The blockchain OffChain worker component then retrieves the OAuth access token from the middleware and submits it as an API POST request to RADAR_OEM_A if Vehicle_OEM_X submits all the certification proofs to the blockchain.
- Fund Transfer including Finalization of Asset Transfer: Then, the bid amount is transferred by the smart contract from escrow to the Vehicle_OEM_X account and

Vehicle_Accounts that participate in this monetization process such that the above condition of the access token and proof are respected.

- Review of transferred Asset: The RADAR_OEM_A that retrieves the processed data from the data pool utilizing the OAuth access token reviews the data. The result of the review process is submitted to the blockchain against the asset ERC721 token issued, which is a fidelity action to augment the price of the data issued by Vehicle_OEM_X in a later transaction or diminish if the review score is bad.

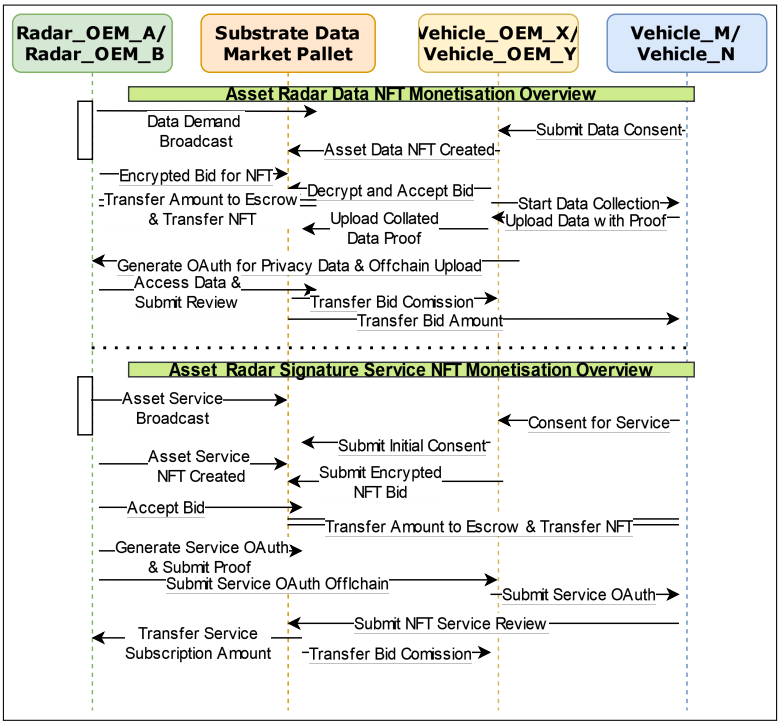

Figure 7. Tokenized Asset and Asset Service Monetization Architecture.

4.2. Tokenized Non-Fungible Asset Data Service Component

The smart contract transaction sequence for the asset service token is represented in Figure 7 for the bidding process similar to the asset component. This is similar to the earlier asset data transaction workflow explained but is in the other sense where the vehicles subscribe to the service offered by RADAR_OEM_A | B. This is explained as follows:

- Road RADAR Signature Asset Service Interest Collation: A particular RADAR_OEM_A has to build the localization, additional and planning layers of additional information over the processed data as the road RADAR signature map service publishes an event to the blockchain smart contract of its availability. This is then processed by Vehicle_OEM_X | Y, where each one gathers the interested vehicle clients for subscribing to this service.
- Road RADAR Signature Asset Service Creation and Bidding: RADAR_OEM_A creates an asset service NFT ERC 721 token and the characteristics of the service with the acceptable price range limit.
- Asset Service Subscription and Proof Generation: Then, each of Vehicle_OEM_X or Y expresses the interest in the subscription service on behalf of its clients as an encrypted bid. RADAR_OEM_A accepts the bid of Vehicle_OEM_X along with its vehicles, transferring funds from the vehicles consented to the escrow account.
- OffChain transfer of Asset Service OAuth API Access Token: RADAR_OEM_A generates an OAuth access token for its road RADAR signature map service along with the hash proof of the service including raw data location, vehicle speed and quality. The OffChain worker retrieves the encrypted OAuth access token from the ledger and submits it to Vehicle_OEM_X.

- Fund Transfer and Finalization of Asset Service Transfer: In parallel, the OffChain worker transfers the funds from the escrow for the bid to RADAR_OEM_A. Also, the asset service token ownership is transferred from RADAR_OEM_A to Vehicle_OEM_X and the interested vehicles.
- Review of Transferred Asset Service: The vehicles or Vehicle_OEM concerned for the asset service token can review the improvement in vehicle driving as a result of road RADAR signature map service, positively or negatively connoting its price and fidelity accordingly.

5. Implementation

This section discusses our decentralized data monetization solution's development and infrastructure deployment operations.

5.1. Substrate Smart Contract Pallet

The Substrate blockchain is an extensible framework allowing for the creation of a client node embedded with smart contracts in the form of a pallet module. We designed this smart contract pallet containing ERC 721 token creation logic, updation, ownership transfer and burn for both asset data and asset service road RADAR signature representation. In our case, the OffChain worker component inside the pallet allows for the smart contract to interface with external API, either as Oracles or token transfers. Fund transfer operation involving escrow and other actors is also defined with logic for proof and data certification submission handled here. The pallet is included in node runtime for the execution of the data monetization application along with the other default functionalities of transaction validation, consensus process and state storage. Our code implementation is realized in 4.0.0-dev with node customizations for testing the BABE and the AuRa consensus algorithm; the code is published publicly.

5.2. Middleware Components

Vehicle_OEM and the RADAR_OEM participate inside the blockchain as an actor in the smart contract and a node validator, respectively. They also participate through OffChain methods by providing the middleware required for collection, processing, transfer, reception and OAuth Access tokens both for raw RADAR data and road RADAR signatures. These business services for RADAR_OEM and Vehicle_OEM are implemented in Java JDK (Java Development Kit) eight language based on the Spring Boot Framework. Its architecture is a REST (Representational State Transfer)-based Model View Controller middleware. It deals with authentification, authorization, OAuth access token generation, verification, management, data storage, retrieval, and other API exposition necessary for the blockchain OffChain worker to interact. It also has the API service to encrypt and decrypt privacy bid payload for asset or asset service, which is Vehicle_OEM or RADAR_OEM. The Unified Modelling Language representation (UML) for the middleware is represented in Figure 8, which comprises the following aspects:

- AuthController: It is the primary class interface in the middleware which accepts the request and redirects it to the necessary business service layer for storing assets, asset service data, OAuth token generation, as well as encryption and decryption process for the bidding procedure.
- UserController: This authentication service retrieves RADAR and RADAR service subscription data.
- Storage: This is the persistence layer for all the details regarding asset, asset service, authentication, and OAuth access tokens.
- Asset/Asset Service Feedback: The asset or asset service feedback in the form of review is also stored OffChain as additional storage.
- RADAR Data: This contains the data payload representation generated by the vehicle and transacted via the blockchain.

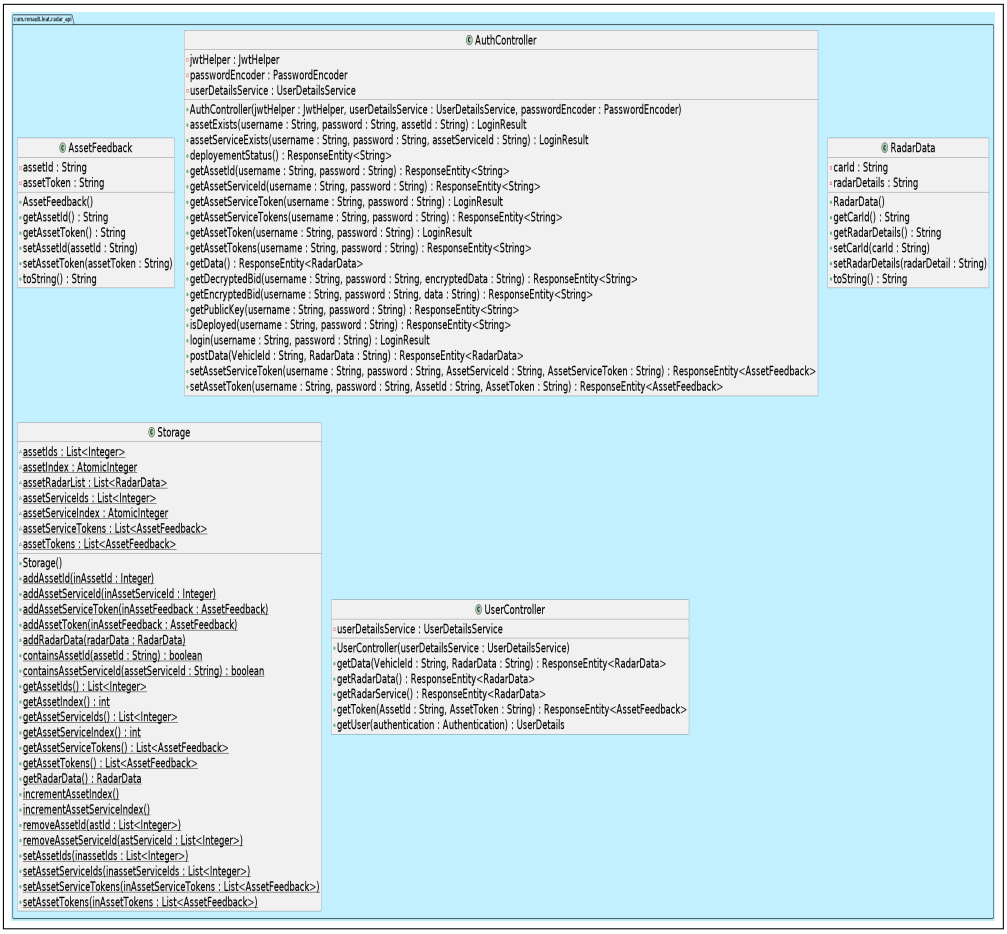

Figure 8. OEM Middleware Data Monetization UML Model.

5.3. Infrastructure

Decentralized data monetization implementation is deployed in the cloud infrastructure represented in Figure 9. As performed earlier for data certification Ethereum-based use-case, we utilize the same TAS cloud infrastructure. The Substrate smart contract pallet containing data monetization logic and the developed middleware are packaged as docker containers. These containers are then deployed on the cloud infrastructure using the Kubernetes orchestration system as pods with necessary extensible data volume, processing, and memory allocations. The client deployed on a separate pod submits a transaction representing Vehicle_OEM or RADAR_OEM to a round-robin-based Nginx load-balancer, which is then redistributed uniformly to the blockchain nodes for validation and execution. The health metrics of the blockchain nodes regarding processor and memory load are stored in the form of the time-series database Prometheus. It is persisted at a pre-determined frequency, then represented graphically in the form of Grafana dashboards deployed in separate pods. The middleware containers of RADAR_OEM and Vehicle_OEM are deployed running on an Apache Tomcat application server. All the internal networking and forwarding among the pods is managed at the Kubernetes level, along with service discovery and load balancing.

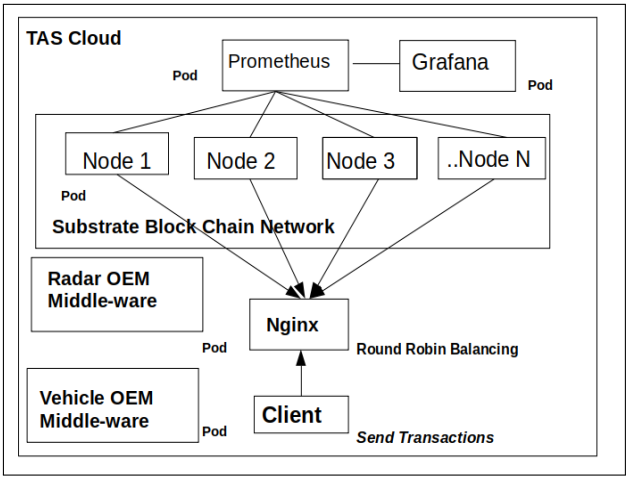

Figure 9. DataMonetization Blockchain Cloud Architecture.

6. Evaluation

In this section, we evaluate the conceptualized data monetization solution from a dual perspective of functional and implementation performance.

6.1. Functional Evaluation

Our virtuous cycle of data monetization since the generation from the vehicle traversing Vehicle_OEM to RADAR_OEM and offering a closed-loop road RADAR signature service based on the received data improvisation has the functional properties as represented in Figure 10. In this discussion, we consider the following representations for the different types of participants in the ecosystem:

- Set of Vehicle_OEM is represented as Vo = $\{Vo_1, Vo_2 \ldots Vo_N\}$ for N participants.
- Set of RADAR_OEM is represented as set of Ro = $\{Ro_1, Ro_2 \ldots Ro_K\}$ for K participants.
- Set of Vehicles are represented as V = $\{V_1, V_2 \ldots V_M\}$ for M participants.
- Processed RADAR data are monetized by each vehicle belonging to a Vo represented as $\{D_1, D_2 \ldots D_L\}$ for L successive data assets to be exchanged over the data market.
- Processed road RADAR signature map (RRS), which a RADAR OEM Ro monetizes, represented as $\{Rs_1, Rs_2 \ldots Rs_P\}$ for P successive RRS asset service to be exchanged as a virtuous cycle in the data market.
- Bid either for asset data or asset data service is represented as $\{B_{x,y,1}, B_{x,y,2} \ldots B_{x,y,S}\}$, where x signifies the bidder and y signifies the asset or asset service tendered.
- Certification for any asset data D_L, C_D is defined as the set of hash signatures derived at each stage of processing steps of individual vehicle data generation, collated raw data in the pool before processing, collated raw data after processing in Vo cloud and then including the OAuth access token.
- Certification for any asset service RSS Rs_P, C_{Rs} is defined as the set of hash signatures derived at each stage of signature formation like initial data state, augmented data after colla ion and improvised RADAR data with signature along with subscription access token.
- Review for any asset or asset service is defined as a score of 0…1 where zero represents the lowest and one represents the highest score; any median value is also possible. This is represented as $R_{i,j}$, where i represents the reviewer and j represents the asset or the asset service transferred and reviewed. Based on the review accumulated for asset data D_L or asset service RSS Rs_P, its future issued asset or asset service can have higher or lower pricing based on its reputation score mean $RSM_{RADAR\_OEM \mid Vehicle\_OEM}$. The reputation score for the asset is proportional to the number of certifications received for the asset or the asset service.

- Reputation Score Mean $RSM_{RADAR\_OEM \mid Vehicle\_OEM}$ for either Vehicle_ or RADAR OEM is the arithmetic mean of historical reviews accumulated. For example, the Reputation score mean for RADAR_OEM or Vehicle_OEM is represented as

$$RM_{\mathrm{j}} = mean(R_{\mathrm{i}-2,\mathrm{j}}, R_{\mathrm{i}-1,\mathrm{j}}, \ldots R_{\mathrm{i,j}})$$

$$RSM_{\mathrm{Radar\_OEM} \mid \mathrm{Vehicle\_OEM}} = mean(RM_{\mathrm{j}-2}, RM_{\mathrm{j}-1}, \ldots RM_{\mathrm{j}})$$

$0 \leq i \leq N$, i represents the index of the reviewers of either asset or asset service

$0 \leq j \leq N$, j represents the index of asset or asset service exchanged

RM_{j}, represents the Mean review attained for a particular asset or asset service j

$RSM_{\mathrm{Radar\_OEM} \mid \mathrm{Vehicle\_OEM}}$ is the Reputation Score Mean of Radar/Vehicle OEM.

Figure 10. Data monetization functional evaluation.

Its properties are discussed below.

- Global Fairness: Fairness in the case of any application is defined by the work of FairSwap [9]: "A fair exchange protocol allows a seller S to sell a digital commodity D for a fixed price p to a buyer B. The protocol is secure if B only pays if he receives the correct D". We extend this work in our above design as in [8] for Global Fairness where every participant in our ecosystem has the following guarantees:
  - Each participant is ensured that fund transfer for any D happens only when the certification conditions is satisfied for C_D.
  - Fund transfer is not fulfilled immediately as it is placed in an escrow account ϵ and then, on verification, transferred to the destined participant account.
  - In either case of the virtuous cycle from RADAR data conversion to road RADAR signature, the facilitators that provision the cloud infrastructure or other value adders can benefit from the commission or fidelity rewards.
  - The asset or asset service token exchange results in the influence of a transparent reputation accumulated by either honest or dishonest activity. This is publicly verifiable and can result in the augmenting or decrease in a reputation participant.
  - To ensure fair pricing, the sealed bid mechanism is encrypted with the bid receiver's public key, which results in a competitive remuneration for either the

asset or asset service. Also, an asset or asset service issuer's reputation organically influences the probability of data or data signature pricing.

- Full Interoperability: Our solution achieves two levels of interoperability, intra-chain and inter-chain, as follows:
  - Intra-Chain: The above monetization solution can integrate with external agnostic (centralized or decentralized) systems through OffChain workers for API requests or responses. Also, as we utilize a balanced mix of events, signed and unsigned transactions (extrinsic) to differentiate the priority of message passing between each actor in the system, unnecessary overhead in the distributed system, especially of consensus, is avoided.
  - Inter-Chain: As it is based on the Substrate framework, this implementation can be extended as a para-chain to integrate with other blockchains or para-chains through Polkadot.
- Chained Data Certification: As mentioned earlier, either for asset RADAR data or asset service road RADAR signature, we ensure the maintenance of the history of the data provenance along with its hash signature-based certification to avoid any counterfeit and validate the health of the data as well as the service.
- Privacy By Design: The solution granularizes the privacy at each level by the following mechanisms:
  - Account Pseudonymization: All the accounts are pseudonymized concerning vehicle participants as it is necessary to identify RADAR_OEMs and Vehicle_OEMs. A vehicle's original identity is another intermediate identity that can be dissociated in case of necessity. This is to respect the forgetting right of GDPR [28] if the vehicle owner decides to remove his information.
  - Bid Sealing: All the bids for either asset or asset service are sealed using encryption, which can be decrypted only by the bid receiver, ensuring competitiveness, transparency, and privacy.
  - Privacy Data Concealment: RADAR data from the origin in the vehicle until the data reception by the RADAR_OEM is privacy treated with the making of sensitive data, and the hash signature generated is added with controlled noise like location, time, etc., to ensure authenticity.
- Dynamic Pricing: As discussed earlier, pricing is based on a sealed bid oriented proportionally to the participant's reputation and review of the exchanged asset as well as asset service data tokens to avoid any price manipulation. This ensures a reputation-based adjustment of the asset pricing, creating a virtuous cycle in fidelity and incentivization.
- Accountable Reputation: Reputation, based on the arithmetic mean of reviewing individual assets transferred as extrinsic transactions diffused through the distributed ledger, is verifiable and transparent for all the ecosystem participants.

6.2. Security

In this section, we analyze our solution for some security concerns based on the amalgamated work of OWASP top ten (Open Worldwide Application Security Project) security threats for web applications and blockchain [29,30]. Also, from the perspective of a smart contract, which, in our case, is an application-oriented blockchain based on a compiled substrate pallet, we analyze its concerns. Their analysis is as follows:

- Injection: The blockchain systems can be compromised by using malicious data, which can be in the form of Integer Overflow or Batch Overflow [30]. We ensure adequate checks and balances in the system where external parameterized signed extrinsic are minimized except for encrypted bids.
- Access Control: Each participant of Vehicle_OEM, RADAR_OEM, Vehicle, and escrow is based on access control ensured through permission pallets and privileged calls as they are consortium-based, limiting risk exposure.

- OffChain Manipulation: Our solution communicates with the OffChain only for retrieval and data job scheduling, not as an oracle for knowledge-based decisions that shield against manipulation attacks.
- Sensitive Data Exposure: Data on the blockchain are only related to the original blockchain accounts and their pseudonymized transactions. All the data are managed OffChain, which ensures that sensitive data are hidden and only managed by the blockchain trust mechanism.
- Smart Contract Security: In this application, our decision of smart contract deployment is not external and is compiled along with the node and executed along with its runtime. Further, the ERC 721 token specification of Ethereum is adapted for the Substrate FRAME library, and each logic is separated for token creation, OffChain data orchestration, as well as certification validation, which is tested for performance.

The functional evaluation was performed by comparison with earlier mentioned works of Ocean Protocol and Digital Infrastructure for Moving Objects (DIMO) in Section 1. We have a consortium where data are exploited to create a mutual benefit among the ecosystem with a measure of balanced privacy, dynamic pricing, fidelity, transparency, certification, and value addition for the vehicle. It has no external dependency in the form of trusted computing or hardware wallets with integrated OffChain seamless interaction for managing and retrieving from external systems.

6.3. Experimental Evaluation

In this section, we experimentally evaluate the monetization architecture implementation as explained earlier in Section 5 deployed in the cloud infrastructure. The study is based on the following dimensions:

- Performance evaluation of the data monetization solution by submitting extrinsic transactions and measuring the throughput of finalized and valid transactions.
- Understanding the hybrid consensus algorithm of AuRa with GRANDPA or BABE with GRANDPA regarding its parameters and deriving an optimum setting, for example, Block Period.
- Analyzing the scalability and fork occurrence in the consensus protocol and evaluating its suitability for enterprise mobility solutions.

6.4. Performance Study

This section discusses the performance results of the data monetization cloud architecture we explained in Section 5.3. The Substrate network is tested with a different configuration of 5, 10, 15, 20 and 25 validator nodes that participate in the variations of (1) AuRa and GRANDPA and (2) BABE and GRANDPA. We organize the test as follows:

- A client that submits the transaction to the load balancer offloads the transaction to a node chosen based on the round-robin algorithm. The client is based on Javascript, which uses the library of Polkadot.js for the following purposes:
    1. Establishing WebSocket communication to the Substrate node and retrieving the meta details of the network like pallet information, account details like a nonce, public key identifier and other miscellaneous details like chain information, fork information, block details and transaction details.
    2. Transaction (extrinsic) construction with the signature using the private key of an account, encoding and decoding transaction payload, submission of the transaction and retrieving its status in the network as finalized or processing.
- For the test, the client submits the transaction of CreateAsset, which creates an asset NFT ERC 721 token in the network. Each transaction has a unique processing complexity based on its purpose logic, affecting its finalization rate or throughput in the network. The test transaction comprises the following complexity:
    1. Calculation of Hash for generating a unique asset Id. Counted as one operation.

2. Storage operation of asset details in asset storage, storing asset count, asset index, and ownership details in separate runtime storage structure either as a storage map, storage double map, or storage value. They are counted as 10 operations.
3. Read operation from storage to retrieve the existing details before any update. They are counted as five operations.

- Transactions (extrinsic) are submitted by the multithreaded client of 1000 threads signed by 2000 pre-generated accounts.
- Then, the transaction is submitted asynchronously at an input rate of 1000 transactions per second with 50,000 total transactions repeated in three iterations to avoid any bias in the result. The block size comprising extrinsic transactions is limited to five megabytes which is optimal as small size can induce message overload and bigger size can delay by a processing bottleneck.
- The submitted signed transactions are checked for finalization, and based on block number, the time difference is calculated for estimating the finalized throughput of the network.

In the next section, we discuss the different test formats across the two consensus algorithm choice variations for the block proposer: AuRa (Authority Round) or BABE (Blind Assignment for Blockchain Extension) and block finalization based on chain level agreement: GRANDPA (GHOST-based Recursive Ancestor-Deriving Prefix Agreement).

6.4.1. Hybrid Consensus Parameter Analysis: Block Period

In this section, we vary the block period in the slot-based block proposition consensus algorithm of either AuRa or BABE. The block authoring or proposition happens at fixed customizable intervals termed the block period. The results of AuRA and BABE block authoring are represented in Figure 11 and Figure 12, respectively. The inference is based on the following explanations:

- Higher interval of block period directly increases the time associated with block production
- Lower or minimal block period decreases the turn-around time for block creation time. Still, it results in a race condition between transaction verification and consensus operation which overflows consensus slot time.

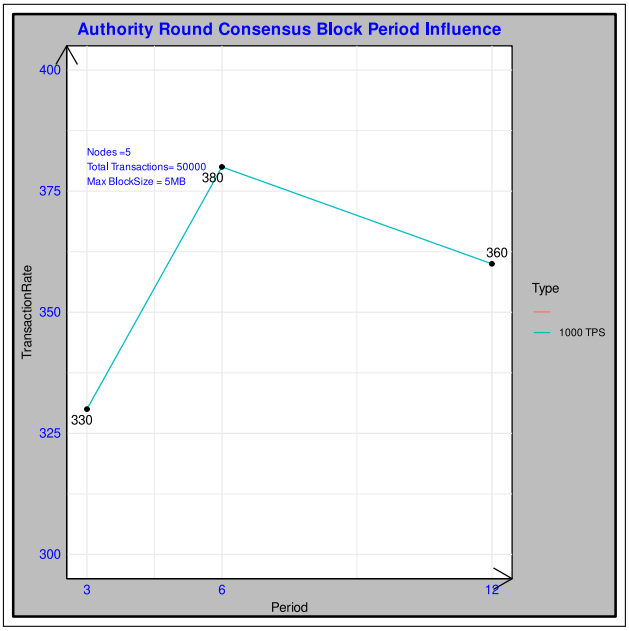

Figure 11. AuRa consensus block period influence.

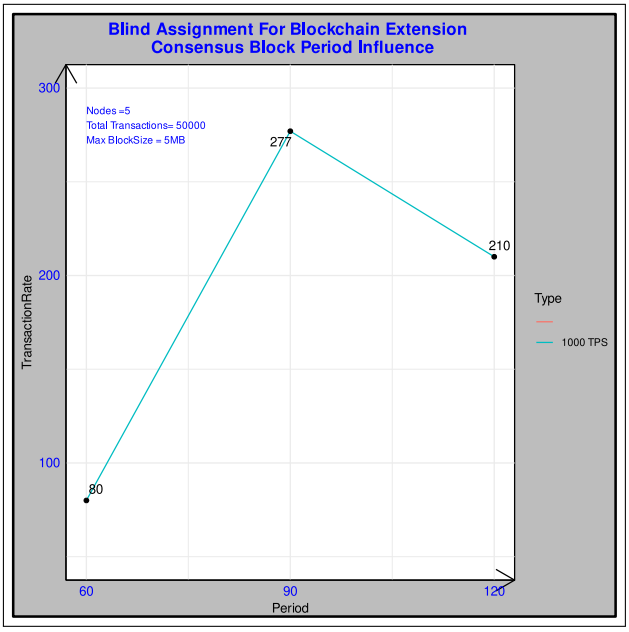

Figure 12. BABE consensus block period influence.

AuRA consensus, as represented in the results, is tested across 3, 6 and 12 s, exhibiting an ideal throughput of 380 transactions per second for a 6 s block period time. The input transactions are around 1000 transactions per second against the output of 380 transactions per second as the transaction processing and consensus factor is to be considered. Meanwhile, the throughput ideal for BABE is around 277 transactions per second. BABE has a high block period time of 60, 90 and 120 s due to the consensus constraints. Smaller block time periods of 3, 6 and 12 s applied for BABE result in a higher amount of forks as a consistency drift is observed in the network. So even though the BABE offers privacy and better security due to its choice of verifiable random computation function-based selection of the proposer, it has a lesser throughput than AuRa due to its computation. Another issue noted here is that there is a GRANDPA-level stalling in the network due to the missing pre-commit votes, which occurs in the case of lesser block periods, which signifies the inconsistency of validators assuring the earlier results experienced in works [31,32].

6.4.2. Hybrid Consensus Scalability Analysis

In this section, we analyze the scalability of AuRa and BABE consensus along with the GRANDPA finalization algorithm. The scalability results for AuRA and BABE are represented in Figures 13 and 14, respectively. The results of AuRa are explained by the $O(n^2)$ message complexity which shows a decreasing throughput in proportion to the number of nodes. In addition, the GRANDPA finalization message complexity bottleneck of $O(n^2)$ further augments the decrease in throughput. Still, the difference in the GRANDPA algorithm is that finalization is performed on the chain of blocks rather than individual blocks of AuRa. Also, another inference based on the result is that higher input transaction per second (TPS) greater than the optimum 1000 increases the forks in the network as well, which in turn decreases the throughput as represented in Figure 15. This fork is explained by the processing and consensus-induced bottleneck affecting the liveness and consistency of chains but resolved by GRANDPA through additional computation.

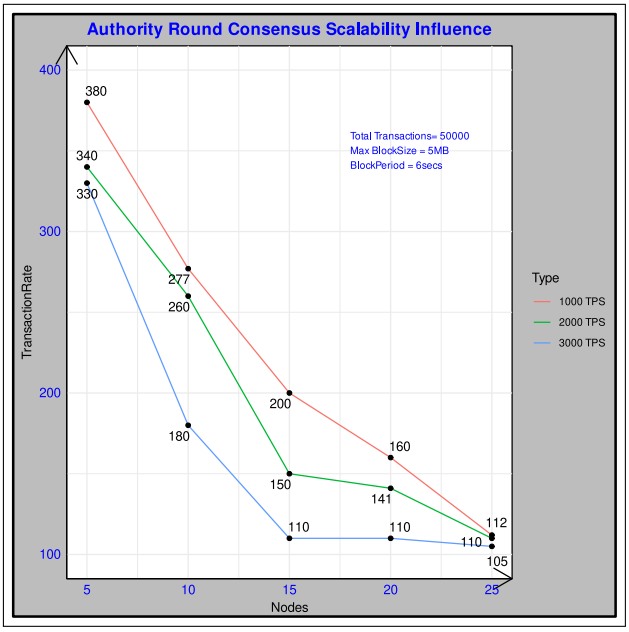

Figure 13. AuRa and GRANDPA consensus scalability performance.

BABE consensus has an algorithmic complexity of O(n) but suffers a decrease in scalability throughput due to the occurrence of forks in the system given the susceptibility of the consensus algorithm to elect multiple validators for a single block height. This results in a larger number of secondary blocks rather than primaries which are attributed to the nature of the algorithm. Accompanied by the computational effort of verifiable random function, it affects the throughput of BABE, which is absent in the AuRa algorithm. The fork occurrence increases with higher input transactions per second, as represented in Figure 16 augmented by increased transaction processing, message communication overhead, and consistency issues.

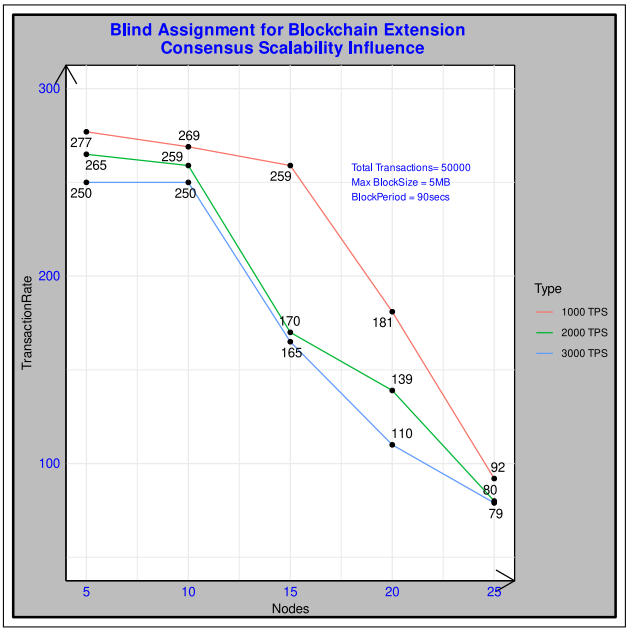

Figure 14. BABE and GRANDPA consensus scalability performance.

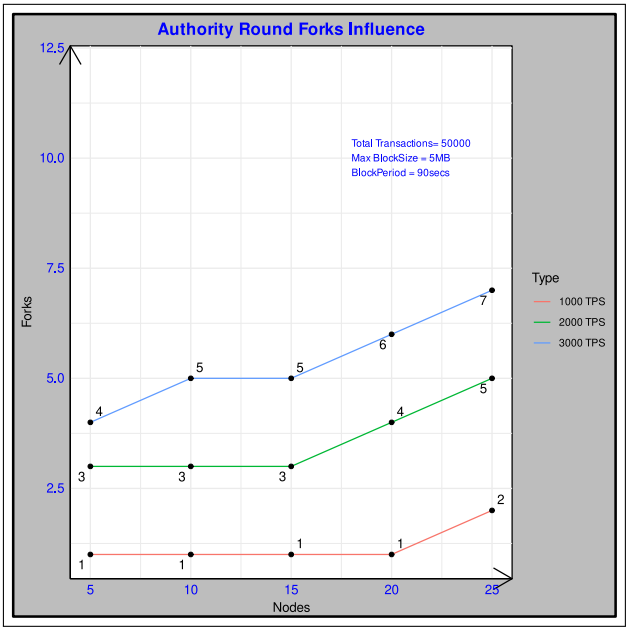

Figure 15. AuRa and GRANDPA consensus fork study.

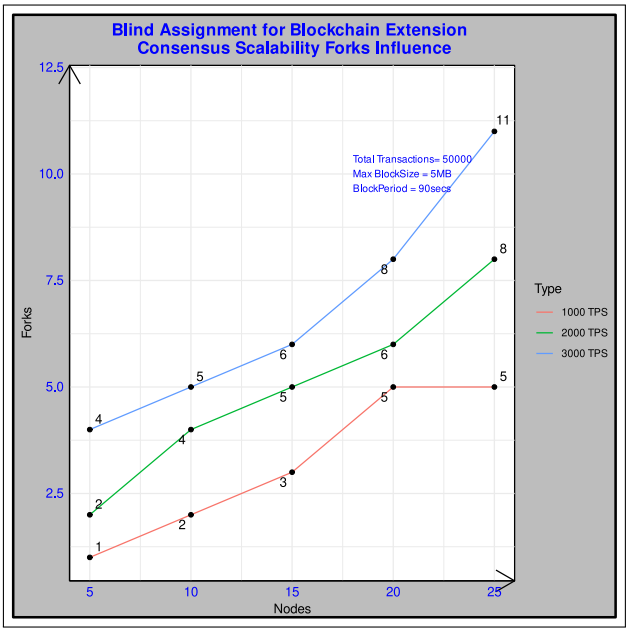

Figure 16. BABE and GRANDPA consensus fork study.

In summary, for the implementation evaluation of data monetization architecture, we realize that the BFT algorithms of AuRa and BABE have a scalability issue. On a positive note, due to their hybrid nature with GRANDPA, they offer finalization and better protocol security but at the cost of additional computation. Both protocols are affected by scalability issues, where the AuRa consensus is more stable than the BABE protocol. BABE, along with the GRANDPA protocol, suffers from consistency issues [31,32] as explained earlier. It offers a higher degree of security in the form of an impossibility of predicting the next successive block proposer using a verifiable random function. GRANDPA BFT offers the finalization of chains in the presence of forks, a hybrid protocol to work with AuRa and BABE in the substrate framework. It is destined to resolve the forks, but its liveness can be affected if most validators are affected by chain consistency issues. The results can be slightly higher if a more straightforward transaction payload is considered. Still, the

throughput of around 300 transactions per second in this data monetization implementation signifies that this architecture is acceptable and can be extended to enterprise mobility use cases.

7. Conclusions

In this work, we analyze the data monetization use case by creating a virtuous cycle of data flow incentivizing Vehicle_OEMs, RADAR_OEMs, and Vehicles who share the data. We exhibit the extensibility of the architecture to external systems and provide RADAR signature services offering global fairness, interoperability, privacy by design, asset data reputation, and a secure network. We further evaluate the cloud-based implementation and understand the Substrate hybrid consensus from an application-based blockchain with an embedded smart contract, including its BFT family consensus scalability limitations.

However, we discuss some improvements that can further enhance the data monetization protocol as below:

- Auction: Auctions can be made more transparent by a commit reveal scheme where there are no time-bound constraints and hidden bids are eventually public when all the bids are received and committed.
- Privacy: Differential privacy [33] can be applied for more granular privacy control on the data ensuring more concealment than the masking techniques. Also, account abstraction based on the ERC 4337 [34] protocol can be considered for anonymizing network participants as an alternative to pseudonymization in our protocol.
- Smart Contract: Smart contract embedded along with client runtime in our Substrate blockchain has better security than by deployment. But it has not been audited using tools like MythX [29], or legal aspects are offloaded, which can be considered in the future.
- Interoperability: Substrate has both intra-chain and inter-chain communication extensibility. The inter-chain exchange in the form of Parachains has several limitations discussed in [35]. They are due to being monetary-based Proof of Stake Polkadot network for parachain slots, validator election transparency issues and governance problems due to the "prime voter" as well as 13 member council governance affecting decentralization. Also, there are liquidity issues with the Polkadot network due to economic reasons, and many parachains like Lido have ceased their operation.
- Certification Proof: Hash-based chained certification can be replaced with Merkle proofs or Zero Knowledge proofs where any user can verify the proof without actual data revealed.
- Consensus: Our implementation analyzes the hybrid consensus of AuRa, BABE, and GRANDPA available in Substrate necessary for our data monetization use case. Another consensus of Nominated Proof of Stake can be tested, available in Substrate Polkadot based on monetary conditions of currency called DOT'S. It can be applied to our use case but needs monetary-based staking and slashing constraints [35].

In evaluating the hybrid consensus algorithms of Clique and BABE with GRANDPA in particular, we notice that there are issues of forks, and scalability impacting the throughput and finalization of the transaction in the blockchain. This behavior is exhibited even in our case of consortium participants with limited strength of nodes. This work, along with our evaluation works on Byzantine Fault-Tolerant (BFT) Algorithms of Practical Byzantine Fault Tolerance (PBFT), Clique, Istanbul Byzantine Fault Tolerance (IBFT), and Quorum Byzantine Fault Tolerance (QBFT) [16–18], confirms the drawbacks of these algorithms. They can be narrowed to scalability, security, forks, data consistency, and deadlock issues.

To solve these problems, we propose an algorithm named CUBA in our other works [36,37] which address these problems. It is a consortium-oriented BFT algorithm where the actions of the participant in terms of block proposal, voting, and round proposal are recorded in the form of utilitarian scores. These scores are then utilized to form the quorum for each round of the consensus process dynamically. Block validation is performed in two phases of intra-quorum and inter-quorum, which is evaluated on a cloud blockchain simulation

network. It offers better performance, security, scalability, and finalization compared to Clique, PBFT, IBFT and QBFT consensus algorithms. This algorithm is considered to be implemented in the substrate framework consensus module and then tested with the same above data monetization use-case implementation offering better blockchain performance.

Author Contributions: Conceptualization, C.N.S. and F.V.; methodology, C.N.S. and F.V.; software, C.N.S.; validation, C.N.S., F.V., S.G. and P.G.-O.; formal analysis, C.N.S.; investigation, C.N.S.; resources, C.N.S. and F.V.; data curation, C.N.S.; writing—original draft preparation, C.N.S.; writing—review and editing, C.N.S. and F.V.; visualization, C.N.S.; supervision, F.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Code Implementation of Data Monetization Substrate pallet along with smart contracts, cloud deployment of Kubernetes, transaction clients, Substrate network configuration files, RADAR and Vehicle Middleware Java implementation as well as test results are released publicly in the GitHub repository: https://github.com/scyrilnaves/these-datamonetisation accessed on 14 January 2024.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Baecker, J.; Engert, M.; Pfaff, M.; Krcmar, H. Business Strategies for Data Monetization: Deriving Insights from Practice. In Proceedings of the 15th International Conference on Wirtschaftsinformatik, Potsdam, Germany, 8–11 March 2020.
2. Ofulue, J.; Benyoucef, M. Data monetization: Insights from a technology-enabled literature review and research agenda. *Manag. Rev. Q.* **2022**, *74*, 521–565. [CrossRef]
3. Network, D. Digital Infrastructure for Moving Objects (DIMO). Available online: https://docs.dimo.zone/docs (accessed on 7 December 2023).
4. Ocean Protocol Foundation. Ocean Protocol: Tools for the Web3 Data Economy. In *Handbook on Blockchain*; Springer: Cham, Switzerland, 2022.
5. Avyukt, A.; Ramachandran, G.; Krishnamachari, B. A Decentralized Review System for Data Marketplaces. In Proceedings of the 2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), Sydney, Australia, 3–6 May 2021; pp. 1–9. [CrossRef]
6. Meijers, J.; Putra, G.D.; Kotsialou, G.; Kanhere, S.S.; Veneris, A. Cost-Effective Blockchain-based IoT Data Marketplaces with a Credit Invariant. In Proceedings of the IEEE International Conference on Blockchain and Cryptocurrency (ICBC), Sydney, Australia, 3–6 May 2021; pp. 1–9. [CrossRef]
7. Al-Sada, B.; Lasla, N.; Abdallah, M. Secure Scalable Blockchain for Sealed-Bid Auction in Energy Trading. In Proceedings of the 2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), Sydney, Australia, 3–6 May 2021; pp. 1–3. [CrossRef]
8. Banerjee, P.; Ruj, S. Blockchain Enabled Data Marketplace—Design and Challenges. *arXiv* **2018**, arXiv:1811.11462. [CrossRef]
9. Dziembowski, S.; Eckey, L.; Faust, S. FairSwap: How to Fairly Exchange Digital Goods. In Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, Toronto, ON, Canada, 15–19 October 2018; pp. 967–984. [CrossRef]
10. FrØland, H.I.; Palm, E.; Kralevska, K.; Gligoroski, D. Web3 and Blockchain for Modernizing the Reseller Market. In Proceedings of the 2023 Fifth International Conference on Blockchain Computing and Applications (BCCA), Kuwait City, Kuwait, 24–26 October 2023; pp. 215–220. [CrossRef]
11. MHP-Riddle & CODE. The Automotive Sector and Blockchain. Available online: https://dlt.mobi/blockchain-and-the-automotive-sector-by-sebastian-becker-and-katarina-preikschat/ (accessed on 9 December 2023).
12. European Telecommunications Standards Institute. Permissioned Distributed Ledger (PDL); Supporting Distributed Data Management. Available online: https://www.etsi.org/technologies/permissioned-distributed-ledgers (accessed on 22 December 2023).
13. International Standards Organization. ISO/TC 307 Blockchain and Distributed Ledger Technologies. Available online: https://www.iso.org/-committee/6266604.html (accessed on 7 November 2023).
14. European Union Blockchain Observatory & Forum. Blockchain Applications in the Automotive Sector. Available online: https://www.eublock-chainforum.eu/news/blockchain-applications-automotive-sector (accessed on 21 November 2023).
15. Continental Automotive. Software-Defined Vehicle. Available online: https://www.continental-automotive.com/en-gl/Passenger-Cars/Technology-Trends/software-defined-vehicles (accessed on 13 January 2024).
16. Samuel, C.N.; Glock, S.; Verdier, F.; Guitton-Ouhamou, P. Choice of Ethereum Clients for Private Blockchain: Assessment from Proof of Authority Perspective. In Proceedings of the 2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), Sydney, Australia, 3–6 May 2021; pp. 1–5. [CrossRef]

17. Samue, C.N.; Severine, G.; David, B.; Verdier, F.; Patricia, G.O. Automotive Data Certification Problem: A View on Effective Blockchain Architectural Solutions. In Proceedings of the 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Virtual, 4–7 November 2020; pp. 0167–0173. [CrossRef]
18. Gerrits, L.; Samuel, C.N.; Kromes, R.; Verdier, F.; Glock, S.; Guitton-Ouhamou, P. Experimental Scalability Study of Consortium Blockchains with BFT Consensus for IoT Automotive Use Case. In Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems, Coimbra, Portugal, 15–17 November 2021; pp. 492–498. [CrossRef]
19. Blair, Tesler (Insight). What Is Privacy by Design and by Default? Available online: https://www.morganlewis.com/pubs/2019/03/the-edata-guide-to-gdpr-what-is-privacy-by-design-and-by-default (accessed on 12 January 2024).
20. What Does Data Protection 'By Design' and 'By Default' Mean? Available online: https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/what-does-data-protection-design-and-default-mean%5Fen (accessed on 17 October 2023).
21. Hameed, K.; Barika, M.; Garg, S.; Amin, M.B.; Kang, B. A taxonomy study on securing Blockchain-based Industrial applications: An overview, application perspectives, requirements, attacks, countermeasures, and open issues. *J. Ind. Inf. Integr.* **2022**, *26*, 100312. [CrossRef]
22. Parity Technologies. Substrate Technology. Available online: https://substrate.io/technology/ (accessed on 8 January 2024).
23. Parity Technologies. Authority Round. Available online: https://paritytech.github.io/substrate/master/sc%consensus%5Faura/index.html (accessed on 20 December 2023).
24. Petrowski, J. Blind Assignment for Blockchain Extension (BABE). Available online: https://polkadot.network/blog/polkadot-consensus-part-3-babe (accessed on 17 December 2023).
25. David, B.; Gaži, P.; Kiayias, A.; Russell, A. Ouroboros Praos: An Adaptively-Secure, Semi-synchronous Proof-of-Stake Blockchain. In *Advances in Cryptology—EUROCRYPT 2018*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 66–98.
26. Stewart, A. Grandpa Finality. Available online: https://research.web3.foundation/en/latest/polkadot/finality.html (accessed on 18 January 2024).
27. Bosch. Road Signature. Available online: https://www.bosch-mobility.com/en/solutions/automated-driving/road-signature (accessed on 22 November 2023).
28. 2018 Reform of EU Data Protection Rules. Available online: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 (accessed on 10 December 2023).
29. Rateb, J. Blockchain Pour L'internet des véHicules: Une Solution IoT Decentralisee Pour la Communication et le Paiement des Vehicules en Utilisant Ethereum. Doctoral Dissertation, HESAM Université, Paris, France, 2021.
30. Jabbar, R.; Kharbeche, M.; Al-Khalifa, K.; Krichen, M.; Barkaoui, K. Blockchain for the Internet of Vehicles: A Decentralized IoT Solution for Vehicles Communication Using Ethereum. *Sensors* **2020**, *20*, 3928. [CrossRef] [PubMed]
31. Stack Exchange. BABE GRANDPA Stalled. Available online: https://substrate.stackexchange.com/questions/214/recovering-from-stalled-finality-babe-grandpa (accessed on 11 December 2023).
32. Wang, Y. The Adversary Capabilities in Practical Byzantine Fault Tolerance. In Proceedings of the Security and Trust Management: 17th International Workshop, STM 2021, Darmstadt, Germany, 8 October 2021; pp. 20–39. [CrossRef]
33. Desfontaines, D.; Pejo, B. SoK: Differential privacies. *Proc. Priv. Enhancing Technol.* **2019**, *2020*, 288–313. [CrossRef]
34. Buterin, V.; Weiss, Y.; Gazso, K.; Patel, N.; Tirosh, D.; Nacson, S.; Hess, T. ERC-4337: Account Abstraction Using Alt Mempool. Available online: https://eips.ethereum.org/EIPS/eip-4337 (accessed on 9 January 2024).
35. Abbas, H.; Caprolu, M.; Di Pietro, R. Analysis of Polkadot: Architecture, Internals, and Contradictions. In Proceedings of the 2022 IEEE International Conference on Blockchain (Blockchain), Espoo, Finland, 22–25 August 2022; pp. 61–70. [CrossRef]
36. Samuel, C.N. Connected Car Communication by DLT Technologies: Mobility Service Implementation by Adaptation of Consortium Blockchain Consensus Algorithms. Doctoral Dissertation, Université Côte d'Azur, Nice, France, 2023.
37. Samuel, C.N.; Verdier, F.; Glock, S.; Guitton-Ouhamou, P. CUBA: An Evolutionary Consortium Oriented Distributed Ledger Byzantine Consensus Algorithm. In Proceedings of the Distributed Computing and Artificial Intelligence, 20th International Conference, Guimaraes, Portugal, 12–14 July 2023; Ossowski, S., Sitek, P., Analide, C., Marreiros, G., Chamoso, P., Rodríguez, S., Eds.; Springer: Cham, Switzerland, 2023; pp. 31–43.