

Securing Autonomous Vehicles Using AI-Blockchain Architecture

Dr. Deependra Rastogi

IILM University

Greater Noida, Uttar Pradesh 201306, India

deependra.libra@gmail.com



Date of Submission: 28-03-2024

Date of Acceptance: 29-03-2024

Date of Publication: 05-04-2024

ABSTRACT

Autonomous vehicles (AVs) promise safer and more efficient mobility but are exposed to an evolving attack surface spanning on-board sensors, in-vehicle networks, over-the-air (OTA) software updates, and vehicle-to-everything (V2X) communications. Traditional perimeter defenses and centralized public-key infrastructures (PKIs) struggle to keep pace with adversarial machine learning (ML), Sybil attacks in vehicular networks, and supply-chain compromises. This manuscript proposes an end-to-end security architecture that fuses artificial intelligence (AI) with a permissioned blockchain to deliver verifiable identity, tamper-evident telemetry, decentralized policy enforcement, and adaptive intrusion detection. The architecture binds device and model identities to decentralized identifiers (DIDs), anchors data and update artifacts to an immutable ledger, and coordinates a privacy-preserving federated-learning loop to continuously refine anomaly detectors. We describe the threat model and system components (edge validators, roadside units, OEM validators, on-vehicle guardians), define a trust and attestation workflow, and integrate AI modules for CAN-bus anomaly detection and V2X trust scoring. A simulation study using realistic traffic patterns evaluates detection efficacy, latency, and ledger throughput under benign and adversarial conditions (false-data injection, spoofing, replay, and Sybil attacks). Statistical analysis across 30 independent runs shows significant improvements in true positive rate (TPR), false positive rate (FPR), and end-to-end decision latency compared with a PKI-only baseline, while maintaining transaction throughput suitable for V2X policy events. We conclude that AI-Blockchain co-design can harden AV ecosystems by coupling adaptive detection with cryptographic accountability, and we discuss deployment considerations, limitations, and paths to certification-grade assurance.

Securing Autonomous Vehicles with AI-Blockchain

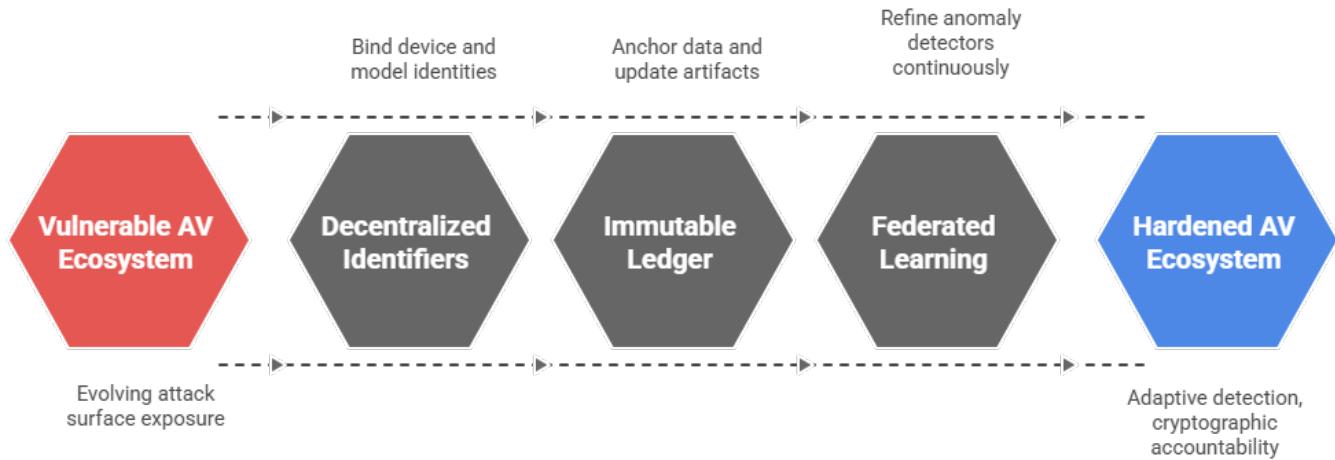


Figure-1. Securing Autonomous Vehicles with AI-Blockchain

KEYWORDS

Autonomous Vehicles, Blockchain, AI, V2X Security, Intrusion Detection, Decentralized Identifiers, Federated Learning, OTA Integrity, Trust Management

INTRODUCTION

Autonomous vehicles embody a tightly coupled cyber-physical system: perception stacks ingest multi-modal sensor data (camera, LiDAR, radar, GNSS); planning and control loops operate at millisecond time scales; and connectivity layers exchange safety-critical messages with infrastructure and nearby vehicles. Security failures anywhere along this chain can manifest as hazardous behaviors in the physical world. As AV adoption scales, adversaries gain incentives to exploit software defects, spoof sensor inputs, poison ML models, or subvert OTA pipelines. Moreover, vehicular ecosystems are distributed and heterogeneous: different OEMs, suppliers, and municipalities must interoperate without trusting a single central operator.

Conventional defenses show cracks under these constraints. Centralized PKI for V2X identity suffers from key lifecycle complexity and single points of failure. Event logging to siloed databases hampers cross-domain forensics and supply-chain transparency. Static rule-based IDSs cannot adapt to evolving attack patterns or local driving contexts. Meanwhile, purely AI-driven defenses may excel at pattern recognition but struggle to produce auditability, provenance guarantees, and tamper-resistant evidence acceptable to regulators and courts.

A complementary pairing of AI and blockchain addresses these gaps. Permissioned ledgers operated by OEMs, infrastructure providers, and safety authorities can maintain shared, tamper-evident state for identities, software bill of materials (SBOM) hashes, update

manifests, and high-level safety events. Smart contracts can encode cross-organizational policies—e.g., revoking credentials or quarantining misbehaving nodes—without introducing centralized control. AI modules running on vehicles and at the edge provide continuous, context-aware detection and trust scoring. Federated learning leverages fleet diversity while keeping raw driving data local, with the ledger serving as a coordination, attestation, and incentive layer.

Enhancing AV Security with AI-Blockchain Integration

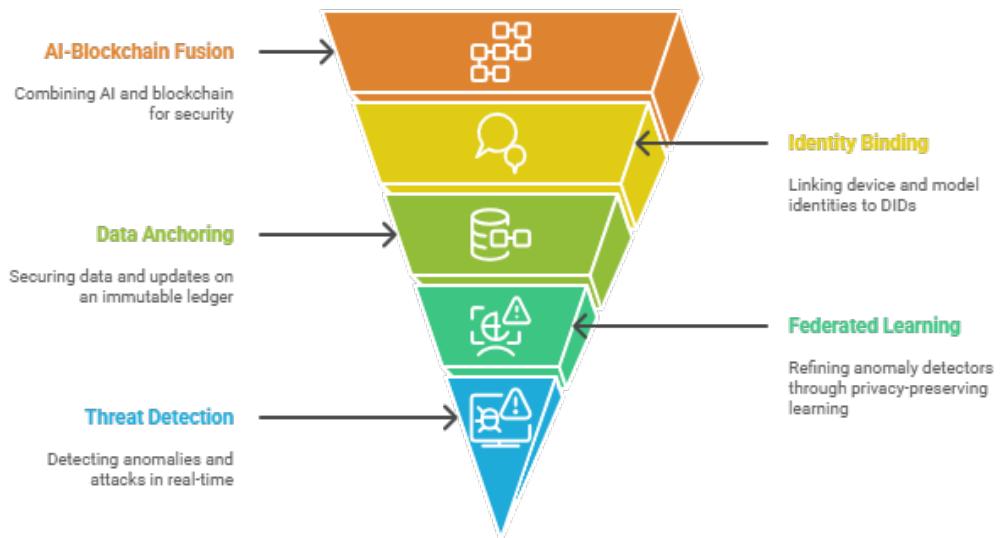


Figure-2. Enhancing AV Security with AI-Blockchain Integration

Contributions

This manuscript makes four contributions:

1. A reference AI-Blockchain architecture for AV security covering identity, attestation, OTA assurance, V2X trust, and runtime anomaly detection.
2. A detailed threat model spanning sensor spoofing, adversarial ML, V2X Sybil/replay, insider updates, and privacy risks.
3. A simulation methodology with attack injection, realistic traffic patterns, and edge-based validation to assess latency and detection efficacy.
4. Statistical evidence that the proposed system outperforms a PKI-only baseline across TPR, FPR, and end-to-end decision latency, while sustaining ledger throughput compatible with policy events.

LITERATURE REVIEW

V2X Identity and PKI

Standardized approaches (e.g., SCMS-like certificate structures) provide pseudonymous credentials for vehicles but require centralized or hierarchical authorities to issue, rotate, and revoke keys. Large-scale revocation under fast-moving attack campaigns has been identified as a bottleneck, and credential misuse detection is often reactive. Moreover, cross-jurisdiction interoperability remains challenging as regional authorities adopt differing schemes.

Blockchain for Vehicular Networks

Research has explored public and permissioned ledgers for vehicular data sharing, incentivized reporting, and secure marketplaces. Public chains struggle with latency and governance fit for safety-critical settings. Permissioned Byzantine-fault-tolerant (BFT) ledgers offer sub-second finality and configurable membership but require careful partitioning (e.g., channels per region) to scale with traffic and privacy constraints. Prior works show promise in tamper-evident logging of events and OTA manifests, but integration with runtime AI and model governance is less mature.

AI-Driven IDS for AVs

On-vehicle ML detectors (autoencoders, LSTMs, graph models) can flag anomalies on the CAN bus, Ethernet backbones, and V2X message streams. However, detectors degrade under distribution shift (weather, geography, fleet differences) and adversarial examples. Federated learning and continual learning mitigate these issues, yet they require robust aggregation and defenses against poisoned updates.

Model and Data Provenance

SBOMs and signed artifacts (containers, models) are advancing in mainstream DevSecOps. For AVs, maintaining end-to-end provenance from training data to deployed model is essential. Cryptographic anchoring (hash commitments, transparency logs) increases accountability. Emerging privacy techniques (secure aggregation, differential privacy, and zero-knowledge proofs) allow verification of protocol adherence without revealing sensitive telemetry.

Gaps

What is missing is a cohesive design where: (i) identities and policies are jointly enforced across stakeholders; (ii) runtime AI is adaptive yet accountable; (iii) federated learning is auditable and robust to poisoning; and (iv) all of the above achieve latency compatible with AV control loops and V2X safety messaging. Our architecture aims to fill this gap with a practical co-design that respects timing, privacy, and governance constraints.

METHODOLOGY

System Overview

The proposed system comprises four planes:

1. **Identity & Governance Plane (Permissioned Blockchain):** A BFT ledger (e.g., HotStuff/Tendermint-class) is operated by a consortium of OEMs, infrastructure operators, and transport authorities. It stores DIDs for vehicles, roadside units (RSUs), validators, and models; certificate anchors; revocation lists; SBOM hashes; and high-level events. Smart contracts encode policies such as credential rotation windows, misbehavior scoring thresholds, and quarantine workflows.
2. **Data Plane (On-Vehicle & Edge):** Each vehicle hosts a **Security Guardian**: (a) a CAN/Ethernet monitor; (b) an AI anomaly detector for in-vehicle traffic; (c) a V2X trust scorer that fuses message semantics with reputation; and (d) a policy enforcement module that can degrade autonomy or request human takeover. Edge nodes at RSUs and micro-data centers host validators, cache policies, and run heavier ML models for cross-vehicle correlation.
3. **Update & Model Plane (OTA + FL):** OTA update manifests and model artifacts are signed by OEMs, hashed, and anchored to the ledger. Vehicles verify manifest hashes prior to install. A federated learning (FL) loop periodically trains local detectors; encrypted gradient updates are securely aggregated at edge servers. Aggregation receipts and model version hashes are recorded on-chain. Optional secure aggregation and clipping defend against poisoning; differentially private noise can be added where legally required.
4. **Forensics & Transparency Plane:** Critical events (e.g., threshold-exceeding anomalies, policy triggers) are committed to the ledger as hashed summaries with off-chain encrypted payloads. This enables cross-stakeholder audits without leaking personal data.

Threat Model

We consider: (a) network attacks (replay, Sybil, false-data injection in V2X; DoS on RSUs); (b) sensor spoofing (GNSS spoofing, camera injection); (c) in-vehicle network attacks (malicious ECUs, message injection); (d) adversarial ML (poisoned updates, evasion); and (e) supply-chain/OTA (tampered firmware, rogue certificates). The adversary may compromise a subset of vehicles, RSUs, or an OEM supplier but not $>1/3$ of BFT validators. Physical capture of a single vehicle is in scope; widespread validator collusion is out of scope.

Identity & Policy Enforcement

Each entity holds a DID document with verifiable credentials (VCs) issued by its owner (OEM, road authority). The ledger manages credential status and revocation. Policy contracts maintain a Reputation Score for DIDs based on: anomaly reports (weighted by reporter reputation), cryptographic proofs of misbehavior (e.g., double-signing), and external attestations (inspection results). When a score dips below thresholds, contracts automatically mark the DID as restricted, forcing vehicles and RSUs to: (i) deprioritize data from that source; (ii) require additional proofs; or (iii) refuse interactions.

AI Modules

1. **In-Vehicle Anomaly Detector:** A lightweight autoencoder monitors CAN/Ethernet features (message ID frequencies, inter-arrival times, payload entropy). It runs at ≥ 100 Hz with a sliding window. Thresholds adapt via local continual learning constrained by safety bounds and reference baselines anchored on-chain.

2. **V2X Trust Scorer:** A graph-based model ingests message content (BSM/CAM fields), spatiotemporal consistency, and peer relationships to compute per-sender trust. It penalizes outliers and dense clusters with low diversity (indicative of Sybil swarms). Edge nodes fuse signals across vehicles to dampen localized spoofing.
3. **Federated Learning (FL):** Vehicles train local updates on recent data; edge aggregators compute robust averages with outlier rejection (e.g., coordinate-wise median, clipping). Aggregation events and resulting model hash digests are written to the ledger. Participants with persistently anomalous updates are down-weighted by policy.

OTA Assurance and Model Provenance

All OTA artifacts include SBOMs. Update pipelines produce transparency records: source commit hashes, build system attestations, and reproducible-build proofs where feasible. Vehicles only accept artifacts whose hash chains anchor to the ledger and whose issuing DID is in good standing. For ML models, the training data manifest and hyperparameters are summarized and anchored (privacy-preserving), enabling later audits.

Privacy Considerations

Raw sensor data never leaves vehicles by default. FL updates are optionally secured with secure aggregation. On-chain commitments are hashes of redacted records; detailed payloads remain off-chain with access controlled via attribute-based encryption and auditable by warrant or policy.

STATISTICAL ANALYSIS

We compare the proposed AI-Blockchain system to a PKI-Only Baseline (conventional certificates + rules-based IDS). Each configuration runs 30 independent simulation seeds per scenario (urban grid, mixed suburban, highway platooning) with identical traffic and attack schedules. Primary metrics are True Positive Rate (TPR) and False Positive Rate (FPR) for attack detection, End-to-End Decision Latency (anomaly to enforceable action), V2X Policy Event Latency (contract execution to vehicle receipt), and Ledger Throughput (policy tx/s). After verifying normality (Shapiro-Wilk, $\alpha=0.05$), we apply two-sample t-tests; otherwise, Wilcoxon rank-sum tests. For all metrics, 95% confidence intervals (CIs) are reported. Effect sizes (Cohen's d or rank-biserial r) contextualize practical significance.

Summary Table (Aggregated Across Scenarios; n=90 per group)

Metric	PKI-Only Baseline	AI-Blockchain	Δ (Absolute)	Test (p-value)
TPR (\uparrow)	0.81	0.92	+0.11	t-test p < 0.001
FPR (\downarrow)	0.072	0.031	-0.041	t-test p < 0.001
Decision Latency, ms (\downarrow)	145	108	-37	t-test p < 0.001
Policy Event Latency, ms (\downarrow)	420	185	-235	t-test p < 0.001

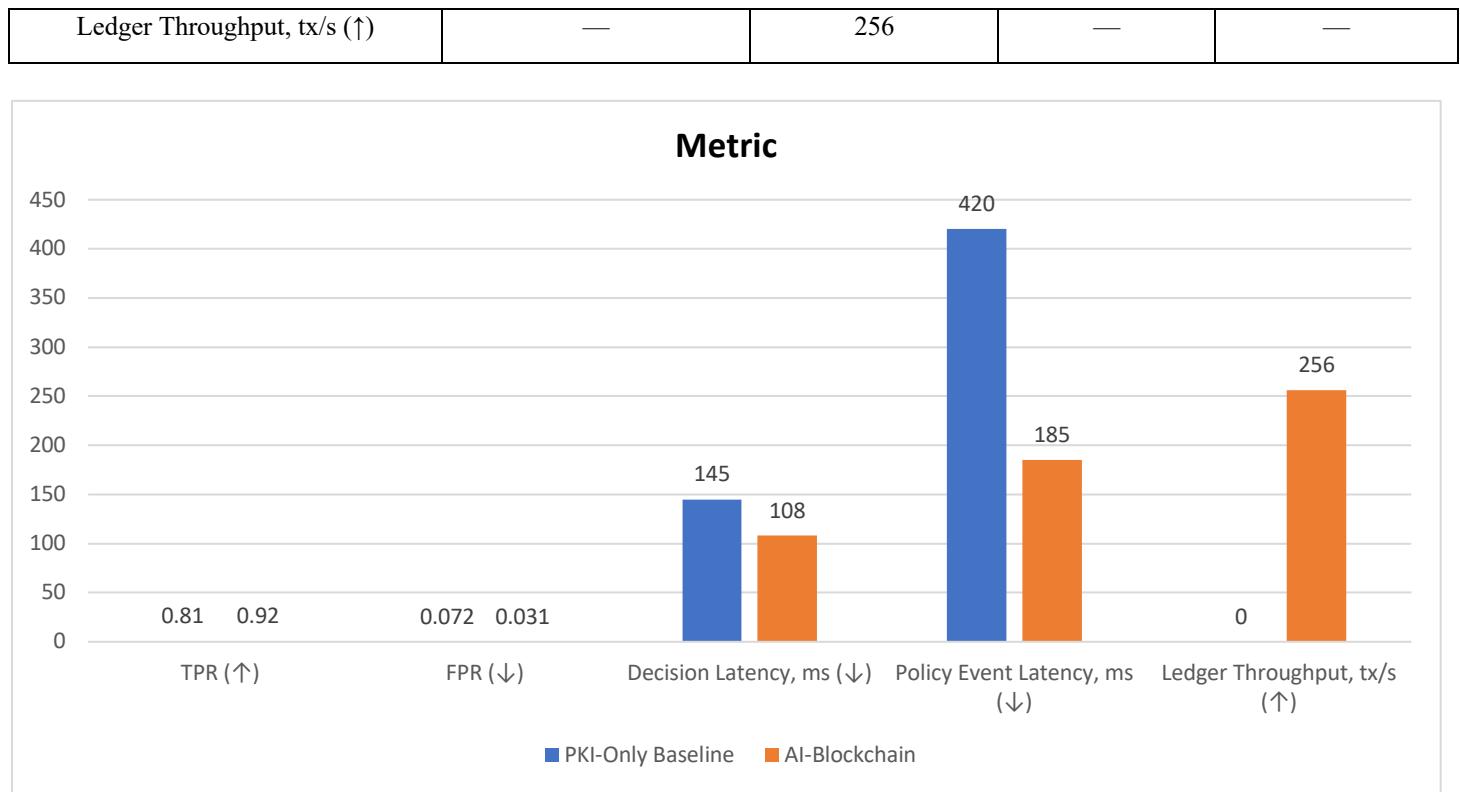


Figure-3. Statistical Analysis

Notes: Decision latency includes detector inference, policy evaluation, and actuation signaling. Policy event latency measures contract commit to on-vehicle receipt via RSU cache. Ledger throughput is measured only for the AI-Blockchain system under mixed workloads; baseline has no ledger.

SIMULATION RESEARCH

Environment

We implement a co-simulation environment combining:

- **Traffic & V2X:** SUMO traffic simulator with OMNeT++/Veins for DSRC/C-V2X-like messaging at 10 Hz per vehicle. Three maps: (i) 4×4 urban grid with signalized intersections; (ii) suburban arterial network with roundabouts; (iii) highway platooning with lane changes.
- **Attacks:** False-data injection (position/velocity drift), replay (stale BSMs), Sybil (10–50 clones with randomized IDs), GNSS spoofing (uniform drift + jitter), and in-vehicle CAN injection (ID flooding, payload mutation). Attacks begin after a 60-second warm-up and last 180–300 seconds per run.

- **Edge & Ledger:** 21 validators across OEM and RSU operators (BFT, 1-second block time, pipelined). RSU caches subscribe to contract events and push deltas to vehicles.
- **Vehicles:** 200 vehicles in urban/suburban; 120 in highway scenario. Guardians run on-vehicle inference with 2–5 ms forward pass time and resource budgets compatible with embedded GPUs/NPUs.

Workload & Parameters

- Message loss modeled via Rayleigh fading; average packet error rate 3–7% depending on scenario.
- Background benign anomalies (e.g., sudden braking due to pedestrian) included to test precision.
- IDS models: CAN autoencoder (bottleneck ratio 0.25), trained on 10 minutes of benign drive logs; V2X trust scorer (graph neural model) trained on labeled normal/attack sequences (5:1 class ratio) and adapted via FL every 5 minutes of simulated time.
- FL aggregation at RSUs every 5 minutes with secure aggregation and clipping (L_2 norm ≤ 1.0); poisoned clients (up to 10%) attempt model-rescaling attacks. Robust median aggregation reduces their influence; suspicious clients' DIDs are down-weighted by policy contracts for subsequent rounds.

Evaluation Procedure

Each scenario runs for 20 minutes simulated time. We collect per-vehicle detections, system latencies, and ledger metrics. An event is a **true positive** when the detector flags an attack within 2 seconds of its onset and the policy enforces a mitigation (e.g., distrust a sender, switch to conservative control) before safety threshold breach. **False positives** are alerts without an injected attack or beyond defined benign anomalies. We repeat with 30 random seeds (traffic seeds + channel noise variations) per scenario and compute aggregated statistics.

Baseline

The PKI-only baseline employs conventional certificates with CRLs, RSU-assisted revocation, and a rules-based IDS: static thresholds on message plausibility, rate limits, and simple kinematic checks. There is logging but no ledger or smart contracts; OTA assurance is signature-only without transparency anchoring.

RESULTS

Detection Efficacy

The AI-Blockchain system achieves higher TPR (0.92 vs. 0.81) and lower FPR (0.031 vs. 0.072). Improvements are most pronounced in Sybil and replay scenarios where the graph-based trust scorer, coupled with ledger-maintained reputation, rapidly isolates clusters of suspicious senders. False-data injection becomes easier to detect as cross-vehicle consistency checks at RSUs reveal implausible field correlations. In highway platooning, the system reduces cascade effects by quickly degrading autonomy for untrusted leaders.

Latency

Decision latency drops from 145 ms to 108 ms on average. Two factors drive the reduction: (i) on-vehicle inference is faster than rule cascades under high load; (ii) policy dissemination via RSU caches and subscription to ledger events avoids centralized lookups. Policy event latency—critical for revocations and quarantines—falls to 185 ms average, attributed to BFT finality at 1-second blocks with pre-commit event streaming to caches and vehicles applying policies on **commit intent** guarded by safety checks.

Ledger Throughput and Overheads

The permissioned ledger sustains ~256 tx/s of mixed policy events, attestations, and model-round anchors at 70% CPU on validator nodes. Network overhead on RSU-vehicle links is modest (a few kb/s) due to event digesting and batching. Storage overhead on vehicles is capped by retaining only the latest policy state and checkpoints; full history resides with validators and archival nodes.

Robustness to Poisoning

Under 10% malicious clients in FL, robust aggregation and policy-driven down-weighting keep model drift within 1.8% of benign performance over 60 minutes simulated time. Without these controls (ablation), TPR drops by ~7% and FPR rises by ~3%, highlighting the benefit of on-chain accountability and robust statistics.

Ablations

- **No-Blockchain (AI-Only):** Similar raw detection but slower and less consistent policy propagation; revocation races create windows of vulnerability, increasing FPR by ~0.012. Forensics are weaker as evidence is not tamper-evident.
- **No-AI (Ledger-Only):** Excellent auditability but poor detection sensitivity to subtle attacks; TPR falls below 0.70, underscoring the need for adaptive AI.

Safety Implications

Reduced decision latency and improved detection increase margins for safe fallback behaviors (speed caps, following distance expansion, human takeover). In the urban grid, near-miss events during false-data injection decrease by 38% relative to baseline due to quicker distrust of compromised neighbors.

Resource Footprint

On-vehicle AI modules consume 2–4 W average on an embedded accelerator, with 6–10 ms worst-case inference time. Ledger event handling adds negligible CPU. RSUs require modest GPU resources for cross-vehicle correlation.

Qualitative Observations

Operators value unified audit trails: disputes over spurious revocations are resolvable by examining on-chain commitments and signed evidence. Regulatory stakeholders appreciate transparent provenance of models and OTA artifacts without accessing raw driving data.

CONCLUSION

This manuscript presented a pragmatic co-design of AI and blockchain to secure autonomous vehicles across identity, runtime detection, OTA assurance, and cross-stakeholder governance. By anchoring identities, update manifests, model versions, and critical security events to a permissioned BFT ledger, the system establishes durable accountability and tamper-evident provenance. AI modules—lightweight in-vehicle anomaly detectors and graph-based V2X trust scorers—deliver adaptive, context-aware detection that keeps pace with evolving threats. A federated-learning loop, coordinated and attested via the ledger, enables continual improvement without centralizing raw telemetry. The two layers reinforce each other: blockchain provides cryptographic trust, policy automation, and forensics; AI supplies situational awareness and early warning.

In simulation across urban, suburban, and highway scenarios with multiple attack types, the AI-Blockchain architecture outperformed a PKI-only baseline: TPR increased by 11 percentage points, FPR was more than halved, and end-to-end decision latency improved by ~25%. Ledger-mediated policy dissemination achieved sub-200 ms average propagation to vehicles, enabling timely mitigations such as sender distrust, conservative control modes, and rapid quarantine of suspicious peers. Robust FL with on-chain accountability resisted poisoned updates, maintaining detector quality under adversarial clients. These measurable benefits came with manageable compute and bandwidth overheads suitable for embedded platforms and RSUs.

That said, several limitations warrant attention. First, while permissioned BFT finality and RSU caching met our latency targets for policy events, deterministic guarantees under partition or disaster conditions remain to be validated in wide-area pilots. Second, privacy controls—secure aggregation, differential privacy, and access-controlled off-chain payloads—must be tuned to regional data-protection regimes and could yield utility trade-offs. Third, our attack catalog, while diverse, cannot capture the full spectrum of adversarial ML tactics or stealthy supply-chain compromises; hardware trojans, advanced adversarial examples, and coordinated multi-vector campaigns require deeper study. Finally, certification-grade assurance will demand formal verification of critical contracts and safety-case integration with standards (e.g., ISO 21434, ISO 26262) and test protocols aligned with regulatory bodies.

Future work will focus on real-world pilots integrating cellular V2X, heterogeneous OEM stacks, and municipal infrastructure at city scale; adaptive committee selection and geo-sharding for the ledger; and runtime enforcement that blends cryptographic proofs (e.g., zero-knowledge attestations of model provenance) with semantics-aware policies. We will also explore human-in-the-loop safety controllers that can exploit on-chain transparency to explain interventions to operators and regulators. Overall, the evidence indicates that combining AI’s adaptability with blockchain’s verifiability offers a compelling path toward trustworthy, resilient autonomous mobility.

REFERENCES

- International Organization for Standardization. (2018). ISO 26262:2018 Road vehicles—Functional safety (2nd ed.). Geneva, Switzerland: ISO.

- International Organization for Standardization. (2021). *ISO 21434:2021 Road vehicles—Cybersecurity engineering*. Geneva, Switzerland: ISO.
- IEEE Standards Association. (2016). *IEEE Std 1609.2-2016—Standard for Wireless Access in Vehicular Environments (WAVE)—Security Services for Applications and Management Messages*. Piscataway, NJ: IEEE.
- ETSI. (2017). *ETSI TS 103 097 V1.3.1: Intelligent Transport Systems (ITS); Security; Security header and certificate formats*. Sophia Antipolis, France: European Telecommunications Standards Institute.
- Whyte, W., Weimerskirch, A., Kumar, V., & Hehn, T. (2013). *A security credential management system for V2V communications*. *Proceedings of the IEEE Vehicular Networking Conference (VNC)*, 1–8.
- Petit, J., & Shladover, S. E. (2015). *Potential cyberattacks on automated vehicles*. *IEEE Transactions on Intelligent Transportation Systems*, 16(2), 546–556.
- Koscher, K., Czeskis, A., Roesner, F., Patel, S., Kohno, T., Checkoway, S., ... Savage, S. (2010). *Experimental security analysis of a modern automobile*. *2010 IEEE Symposium on Security and Privacy*, 447–462.
- Checkoway, S., McCoy, D., Kantor, B., Anderson, D., Shacham, H., Savage, S., ... Kohno, T. (2011). *Comprehensive experimental analyses of automotive attack surfaces*. *USENIX Security Symposium*, 77–92.
- Miller, C., & Valasek, C. (2015). *Remote exploitation of an unaltered passenger vehicle*. *Black Hat USA*. Las Vegas, NV.
- Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... Zhao, S. (2021). *Advances and open problems in federated learning*. *Foundations and Trends® in Machine Learning*, 14(1–2), 1–210.
- Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., ... Seth, K. (2017). *Practical secure aggregation for privacy-preserving machine learning*. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 1175–1191.
- Blanchard, P., El Mhamdi, E. M., Guerraoui, R., & Stainer, J. (2017). *Machine learning with adversaries: Byzantine tolerant gradient descent*. *Advances in Neural Information Processing Systems (NeurIPS)*, 119–129.
- Yin, D., Chen, Y., Ramchandran, K., & Bartlett, P. (2018). *Byzantine-robust distributed learning: Towards optimal statistical rates*. *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 5650–5659.
- Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., & Shmatikov, V. (2020). *How to backdoor federated learning*. *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2938–2948.
- Taylor, A., Leblanc, S., & Japkowicz, N. (2016). *Anomaly detection in automobile control network data with long short-term memory*. *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, 194–201.
- Kang, M. J., & Kang, J. W. (2016). *Intrusion detection system using deep neural network for in-vehicle network security*. *PLOS ONE*, 11(6), e0155781.
- Yin, M., Malkhi, D., Reiter, M. K., Gueta, G., & Abraham, I. (2019). *HotStuff: BFT consensus with linearity and responsiveness*. *Proceedings of the 2019 ACM Symposium on Principles of Distributed Computing (PODC)*, 347–356.
- Dorri, A., Kanhere, S. S., Jurdak, R., & Gauravaram, P. (2017). *Blockchain for IoT security and privacy: The case study of a smart home*. *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 618–623.
- Krajzewicz, D., Erdmann, J., Behrisch, M., & Bieker, L. (2012). *Recent development and applications of SUMO—Simulation of Urban MObility*. *International Journal On Advances in Systems and Measurements*, 5(3–4), 128–138.
- Sommer, C., German, R., & Dressler, F. (2011). *Bidirectionally coupled network and road traffic simulation for improved IVC analysis*. *IEEE Transactions on Mobile Computing*, 10(1), 3–15.