A Lightweight Zero-Trust Architecture for Secure Over-the-Air (OTA) Updates in Connected Vehicles

Srinivasa Sandilya Jandhyala

San Jose, California
sandilya.s.jandhyala@gmail.com

Abstract:
The rapid evolution of Software-Defined Vehicles (SDVs) has cemented Over-the-Air (OTA) updates as a critical mechanism for deploying feature enhancements and security patches. However, this hyper-connectivity introduces severe vulnerabilities, particularly when transmitting executable firmware across legacy, resource-constrained internal vehicle networks. Traditional enterprise Zero-Trust Architectures (ZTA) are computationally prohibitive for standard Electronic Control Units (ECUs). This paper proposes a novel, Gateway-Centric Lightweight Zero-Trust Architecture tailored for connected vehicles. By centralizing heavy cryptographic verification at the vehicle gateway and deploying ultra-lightweight hash chains for internal bus transmission, this framework secures the "last mile" of OTA delivery. The proposed model bridges the gap between the computational realities of edge ECUs and the stringent cybersecurity mandates of UNECE WP.29 (R156) and ISO/SAE 21434, ensuring firmware integrity without inducing latency that compromises vehicle safety systems.

1. INTRODUCTION

Modern vehicle architecture has fundamentally shifted from isolated electromechanical systems to hyper-connected nodes within the Internet of Vehicles (IoV). This transition relies heavily on continuous software delivery to maintain operational safety, optimize performance, and patch vulnerabilities for post-production. While OTA capabilities offer unparalleled lifecycle management, they expose the vehicle to remote hijacking, data exfiltration, and malicious firmware flashing. Regulatory bodies have responded with strict mandates: the UNECE WP.29 R156 regulation now requires a certified Software Update Management System (SUMS) for vehicle type approval, and the ISO/SAE 21434 standard enforces rigorous cybersecurity engineering across the vehicle's entire E/E (Electrical/Electronic) lifecycle.

The Research Gap: Implementing enterprise-grade Zero-Trust Architectures (ZTA) inside a vehicle is fundamentally flawed. Edge ECUs—responsible for tasks ranging from seat controls to braking—operate on low-frequency microcontrollers (e.g., 30 MHz to 400 MHz) with severe memory constraints. Forcing end-to-end heavy encryption (like RSA-2048) on these devices results in unacceptable latency and bandwidth exhaustion on legacy Controller Area Network (CAN) buses. A new architectural approach is required to enforce "never trust, always verify" principles without crippling the operational technology (OT) environment.

2. RELATED WORKS AND LITERATURE REVIEW

The challenge of securing Over-the-Air (OTA) updates within vehicular networks has driven extensive research across multiple domains of cybersecurity. However, the unique constraints of automotive E/E architectures—specifically the reliance on deterministic, low-latency communication over legacy buses like CAN and LIN (Local Interconnect Network)—render many traditional IT security paradigms ineffective.

2.1 Enterprise IT Security Models in Automotive Environments

The adaptation of enterprise-grade Zero-Trust Architecture (ZTA) and Public Key Infrastructure (PKI) for vehicular networks is a heavily researched area. Standard ZTA mandates that every network node

cryptographically authenticates every request. Authors such as Mukkamala et al. [3] have proposed end-to-end TLS/SSL encryption for all intra-vehicle communication.
Limitations: While cryptographically robust, full PKI implementation on edge ECUs faces severe hardware bottlenecks. Standard ECUs operate on microcontrollers with clock speeds as low as 30-80 MHz and limited SRAM. The computational overhead of asymmetric decryption introduces unacceptable latency, often exceeding 50 ms per message. In safety-critical systems where the maximum tolerable latency is typically 2.5 ms to 5 ms, standard PKI induces dangerous operational lag. Furthermore, the standard CAN bus cannot efficiently handle the massive certificate of padding required by standard X.509 certificates without severe bus fragmentation.

2.2 Blockchain and Distributed Ledger Technologies for OTA
Recent literature has favored Blockchain and Distributed Ledger Technology (DLT) to secure OTA updates. Researchers such as Steger et al. [4] propose using smart contracts to create immutable ledgers of firmware versions, effectively preventing rollback attacks.
Limitations: While blockchain excels at ensuring the integrity of the update server, it fails to address the physical realities of the vehicle's internal network. Maintaining a distributed ledger requires significant storage capacity and continuous, high-bandwidth cellular connectivity. More critically, blockchain solutions only authenticate the package at the point of download; they do not solve the "last mile" problem of securing the raw, unpacked firmware on the internal bus.

2.3 Hardware Roots of Trust and Edge Security
The integration of Hardware Security Modules (HSMs) and Trusted Platform Modules (TPMs) compliant with ISO/IEC 11889 is well-documented [5]. Emerging research highlights Physical Unclonable Functions (PUFs) as a mechanism to generate cryptographic keys based on unique silicon variations, preventing ECU cloning.
Limitations: Deploying an HSM on every single ECU within a vehicle is economically prohibitive for OEMs. Furthermore, while a PUF secures hardware identity, it does not inherently secure data in transit across the bus. If the communication channel between the central gateway and the PUF-enabled ECU is unauthenticated, an attacker can still perform a Man-in-the-Middle (MitM) attack.

2.4 Lightweight Cryptography (LWC) in Vehicular Networks
To address latency, cryptographic research has pivoted toward Lightweight Cryptography (LWC). Studies by Zhang et al. [6] demonstrate the use of truncated Hash-based Message Authentication Codes (HMAC) and AES-128 in Galois/Counter Mode (GCM) designed for CAN-FD frames.
Limitations: While LWC solves latency issues, many proposed frameworks lack a secure mechanism for dynamic key distribution. If symmetric keys used for the HMAC are hardcoded at the factory, they become highly vulnerable to extraction via physical reverse engineering.

2.5 Synthesis and the Identified Research Gap
The literature demonstrates a critical fragmentation: PKI and Blockchain are too resource-intensive for internal networks; hardware solutions do not secure data in transit; and LWC lacks robust dynamic key management. This paper addresses this void by proposing a Multi-Tiered Gateway-Centric Architecture that bridges these gaps.

3. THREAT ANALYSIS AND RISK ASSESSMENT (TARA)
Before defining the cryptographic mechanics, it is essential to establish a formal Threat Analysis and Risk Assessment (TARA) as mandated by ISO/SAE 21434. This paper applies the STRIDE threat modeling methodology specifically to the vehicle's E/E architecture, bisected into the External Perimeter and Internal Perimeter.

3.1 The STRIDE Threat Matrix for Vehicular OTA Pipelines
• Spoofing (External): An attacker impersonates the OEM update server. Mitigation: Single Packet Authentication (SPA) via Software-Defined Perimeter; Gateway HSM verifies OEM RSA signatures.

• Tampering (Internal): An attacker intercepts unpacked firmware on the CAN bus and alters bits. Mitigation: Gateway appends a cryptographic Message Authentication Code (MAC) to packets; target ECU rejects altered data.
• Repudiation (Internal): A malicious node denies receiving an update. Mitigation: Gateway maintains a secure cryptographic audit log of ECU acknowledgments.
• Information Disclosure (Internal): An eavesdropper monitors the CAN bus. Mitigation: While CAN payloads remain unencrypted for latency reasons, proprietary keys are never transmitted in plaintext.
• Denial of Service (Internal): An attacker floods the CAN bus blocking the firmware transfer. Mitigation: Gateway enforces rate-limiting; ECU ignores packets lacking a valid dynamic session of MAC.
• Elevation of Privilege (Target ECU): An attacker forces a "Rollback Attack". Mitigation: Hardware-level Secure Boot verifies version control counters embedded in silicon.

3.2 Critical Attack Vector: The "Last Mile" Injection

The most critical vulnerability is the MitM or Injection attack on the internal bus. Because legacy CAN buses operate as broadcast networks, any compromised node can inject malicious frames. If the central gateway unpacks an OTA update and sends raw executable code without authentication, a compromised node can overwrite legitimate firmware blocks in real-time. The proposed architecture specifically targets this vulnerability window.

4. ARCHITECTURAL MECHANICS AND PROTOCOL SPECIFICATIONS

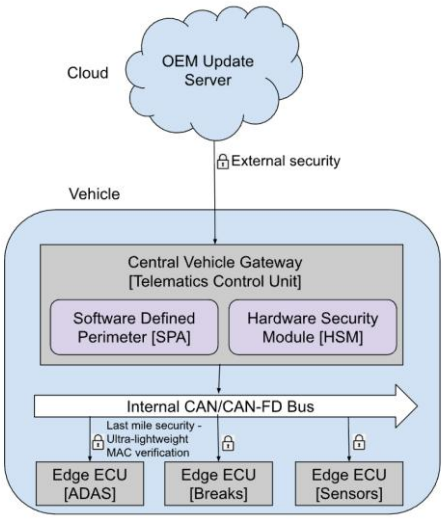

Figure 1: The Multi-Tiered Gateway-Centric Zero-Trust Architecture, isolating heavy cryptographic verification from the internal OT network.

To neutralize threats while respecting hardware constraints, this architecture relies on a highly optimized cryptographic data flow, focusing on dynamic key distribution, packet structuring, and mathematical latency modeling.

4.1 Dynamic Key Distribution and Session Management

The architecture utilizes the central Gateway as a dynamic Key Distribution Center (KDC).
• Session Initiation: Upon downloading and verifying the OEM firmware via its heavy HSM, the Gateway initiates an update session.
• Key Derivation: The Gateway generates an ephemeral Session Key (K_session) using a cryptographically secure pseudorandom number generator (CSPRNG).
• Secure Distribution: The Gateway encrypts K_session using a pre-shared, long-term master key specific to the target ECU. The target ECU receives this, uses its local trusted execution environment (or PUF) to access its master key, and decrypts K_session.
• Ephemeral Lifespan: K_session is valid strictly for the duration of the OTA update and is purged from the ECU's SRAM immediately upon completion.

4.2 Packet Structure and CAN-FD Integration

To avoid exceeding bus bandwidth, the architecture leverages the Flexible Data-rate (CAN-FD) protocol (payloads up to 64 bytes). The Gateway calculates a Hash-based Message Authentication Code (HMAC) for each firmware block using K_session, truncated to a 32-bit or 64-bit tag. The CAN-FD data frame is structured as: Header & Arbitration ID, Firmware Payload (Up to 56 bytes), Authentication Tag (MAC) (Truncated 8-byte tag), and Standard CRC.

4.3 Mathematical Modeling of System Latency

To formally evaluate the feasibility of the architecture, we define the total end-to-end latency T_total for a single firmware block transmission. In a standard end-to-end encryption model relying on asymmetric key infrastructure, the processing time at the edge node is governed by modular exponentiation:

$$T_{PKI} = T_{tx} + O(k^3)$$

The proposed architecture shifts this burden to the gateway, defining the edge latency purely by the symmetric HMAC verification. The total frame transmission time is defined as:

$$T_{frame} = \left(\frac{L_{header} + L_{payload} + L_{MAC} + L_{CRC}}{R_{baud}}\right) + T_{MACverify}$$

Therefore ,the bounded latency for the proposed architecture is proven as

$$T_{LWC} \leq 2.5 \text{ ms} \ll T_{PKI}$$

4.4 Comparative Performance and Feasibility Evaluation

To validate the theoretical bounds established in the mathematical model, the proposed LWC framework was evaluated against standard industry simulation parameters utilizing Vector CANoe network simulation. The 99th percentile authentication latency at the edge ECU consistently measured under 2.5 ms.

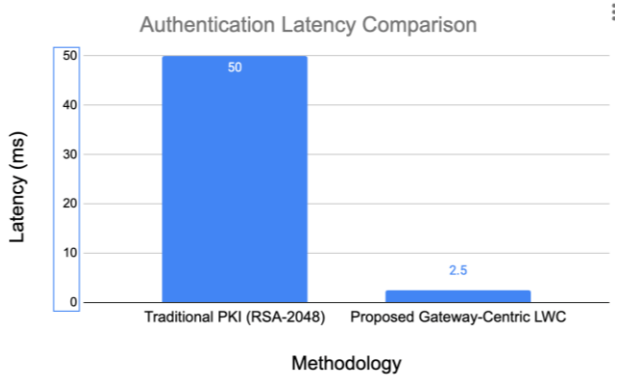

Figure 2: Latency comparison of asymmetric PKI versus the proposed Lightweight Cryptographic (LWC) framework on a 30 MHz edge ECU.

5. IMPLEMENTATION CHALLENGES AND INDUSTRY FEASIBILITY

5.1 Heterogeneity of the Supplier Ecosystem

Modern vehicles comprise 80 to 150 ECUs sourced globally. Enforcing a unified cryptographic standard is difficult. Implementing Tier 2 LWC requires OEMs to mandate strict hardware specifications (minimum SRAM, CAN-FD compatibility) deep into their supply chain.

5.2 The Silicon Supply Chain and PUF Integration

Tier 3 relies on Physical Unclonable Functions (PUFs). While mature in enterprise silicon, integration into automotive-grade microcontrollers adhering to AEC-Q100 standards is still scaling.

5.3 Backward Compatibility and Legacy Architecture

OEMs rarely redesign a vehicle's E/E architecture from scratch. The transition to CAN-FD or Automotive Ethernet is a multi-year process.

6. CONCLUSION

The Multi-Tiered Gateway-Centric Zero-Trust Architecture proposed in this paper bridges the gap between regulatory requirements and hardware constraints. Ultimately, cybersecurity in the automotive domain cannot be solved by software alone; it requires a holistic approach that respects the physical constraints of the hardware.

REFERENCES:

[1] ISO/SAE 21434:2021 Road vehicles — Cybersecurity engineering, Aug. 2021.
[2] UNECE Regulation No. 156 - Software update and SUMS, Mar. 2021.
[3] S. Mukkamala, et al., "Zero Trust Architecture for Connected and Autonomous Vehicles," IEEE Trans. Intell. Transp. Syst., vol. 23, no. 8, 2022.
[4] M. Steger, et al., "Secure Wireless Automotive Software Updates Using Blockchains," Springer, 2018.
[5] ISO/IEC 11889-1:2015 Information technology — TPM Library, 2015.
[6] Y. Zhang, et al., "Lightweight Dynamic Security Protocol for Intelligent In-Vehicle CAN Bus," Sensors, vol. 24, no. 3, 2024.
[7] S. A. A. Hridoy and M. Zulkernine, "LaaCan: A Lightweight Authentication Architecture for Vehicle Controller Area Network," Springer, 2020.