Poster: Privacy Implications of Personally Identifiable Information in Discarded Vehicle Entertainment Systems: An Analysis in the Era of Large Language Models

Bhargab Acharya
University of Memphis

Christos Papadopoulos
University of Memphis

Spiros Thanasoulas
University of Memphis

Sam Lauzon

Bidhya Shrestha
University of Memphis

1 Introduction

Modern vehicles have evolved into software-defined platforms with hundreds of Electronic Control Units (ECUs), as mechanical features are increasingly being replaced by software systems. These vehicles exhibit deep network connectivity through multiple layers order to provide diverse features ranging from personalized multimedia settings, to assisted driving. It is connected internally via automotive networks like CAN bus and automotive ethernet connecting the numerous ECUs, and externally through telematics systems, internet connectivity, and integration with personal devices, collectively generating substantial repositories of personal data. However, they face similar challenges to IoT devices: fragmented update mechanisms, lack of standardized security frameworks, and extended deployment lifespans.

A recent Mozilla study [2] characterized the state of vehicle privacy as "a privacy nightmare on wheels," with manufacturers collecting extensive data while providing limited control to users. Our research demonstrates how even low-sensitivity information from discarded vehicle infotainment systems can be leveraged for sophisticated social engineering attacks when combined with Large Language Model (LLM) inference capabilities.

The scale of this privacy risk is significant: U.S. Census data shows over 150 million light-duty vehicles from 2014-2016 remain on the road [3]. With the vehicle salvage industry processing millions of vehicles annually, the exposure surface is substantial. Our work documents personal data persistence in modern vehicle infotainment systems; provides a systematic, low-cost methodology for data extraction from salvaged automotive systems; reveals outdated software ecosystems in recent model vehicles; and demonstrates how LLM-enhanced inference capabilities transform low-sensitivity data into high-sensitivity insights.

2 Methodology

2.1 Hardware Acquisition

Our research began with acquiring used vehicle entertainment systems through secondary marketplace: (eBay, $20-100 per unit). We acquired diverse test units spanning 2011-2020 models across multiple manufacturers (BMW, Jeep, Honda), demonstrating the widespread availability of used automotive components containing personal data.

2.2 Data Extraction Methods

For hard drive-based systems (BMW, Jeep), we used physical disassembly with basic tools (Torx T8 security screws) to remove standard 80GB IDE 2.5" hard drives. These were connected to a forensic workstation via IDE-to-USB adapter ($20), creating disk images using the `dd` utility, followed by partition extraction and analysis via current best known QNX parsing scripts [1]. For MMC-based systems (Honda), we accessed the diagnostic menu through button combinations (audio + menu + brightness), connected to WiFi through developer settings, and exploited a common and old linux vulnerability for root access to create a complete disk image. Our findings reveal that the 2016 Honda Pilot ran Android Jelly Bean (released 2012), demonstrating how even recent vehicles often run severely outdated software with known vulnerabilities.

2.3 Data Analysis Approaches

Our analysis mainly involved two techniques. Pattern searching in the binary image and file search in a mounted disk. Simple `strings` with `grep` was enough to reveal lots of sensitive informations like GPS locations, contacts, browser history etc. File system analysis involved a modified QNX6 parser script for BMW/Jeep systems and standard Linux mounting for the Honda Android system, enabling database analysis (SQLite) for structured data and examination of various log

files. Advanced forensics utilized `Foremost` and `Autopsy` for deleted file recovery and custom regex patterns for identifying sensitive information.

3 Key Findings

Our analysis revealed extensive personal information across all systems. Contact information included names, phone numbers, email addresses, and Bluetooth device pairings. Location data provided GPS coordinates, frequent destinations, and navigation history. Account credentials included HondaLink authentication token and Aha app credentials. Usage patterns were found in detailed power logs, revealing consistent driving schedules and habits. Vehicle data included complete VINs, service history, and diagnostic information. Multimedia consisted of recoverable photos of previous owners, audio files, and browser history.

Most concerning was the HondaLink account. If the user used the same account in their new Honda vehicle, we could remotely start/stop engine, unlock doors, track location, and access to maintenance information. This demonstrates how salvaged components can potentially compromise not just data privacy but also physical security of vehicles still in operation.

4 LLM-Enhanced Attack Pipeline

Even "low-sensitivity" data points can be transformed into highly sensitive inferences through modern LLMs. We developed a three-stage attack pipeline using locally hosted LLMs where (1) Infer sensitive attributes not explicitly present in the original data; (2) Identify optimal attack vectors; and (3) Produced personalized phishing content leveraging both explicit data points and inferred attributes.

Our LLM pipeline successfully transformed seemingly innocuous data fragments into detailed user profiles with minimal technical expertise. Contact names with slang terms allowed demographic inference. Location data enabled precise identification of home and frequent destinations. Contact lists revealed relationships and social connections. Using these inferences, our system generated highly personalized phishing content across multiple attack vectors, even with distilled LLMs (LLama:70B parameters) generating convincing content within few minutes on consumer hardware.

5 Implications and Recommendations

The combination of outdated software platforms in modern vehicles, persistent storage of sensitive user data, and the inference capabilities of modern LLMs creates a significant privacy and security risk that challenges traditional privacy frameworks. Unlike IoT devices, vehicles have extended lifecycles, with census data showing approximately 12-13 million light-duty vehicles sold annually, many remaining in use for 15+ years.

The issue resembles the IoT security challenge but at a larger scale: a 2016 Honda Pilot running Android Jelly Bean (equivalent to iPhone 6) demonstrates how even relatively recent vehicles often run severely outdated software with known vulnerabilities. Our recommendations include: manufacturer implementation of mandatory secure deletion mechanisms during factory resets; revision of automotive privacy frameworks to account for LLM inference capabilities; regular security updates throughout the vehicle's lifecycle, not just during active production; industry-wide adoption of privacy-by-design principles for vehicle software architectures; and consumer education on proper data sanitization before vehicle/device disposal.

6 Conclusion

This work exposes a critical privacy and security vulnerability at the intersection of automotive technology and advanced AI systems. Modern vehicles (2012+) behave like software-defined platforms but lack security practices like software platforms. A 2016 Honda Pilot runs Android Jelly Bean (2012)—equivalent to iPhone 6 era software—with known exploitable vulnerabilities. Sensitive user data persists in vehicle systems even after disposal. Low-cost, readily available tools and techniques can extract this data. Any vehicle with a hard drive/SSD is subject to same attacks if storage is not encrypted. Modern LLMs can transform seemingly innocuous data into highly sensitive inferences, enabling personalized social engineering attacks.

Unlike smartphones or computers, vehicles often remain in service for 15+ years with minimal software updates, creating an expanding universe of vulnerable platforms containing personal data. With millions of vehicles changing hands annually in the U.S. alone, our findings urge immediate attention to this emerging privacy threat vector that will only grow more significant as LLM capabilities advance.

Data source: U.S. Census Bureau, Vehicle Inventory and Use Survey, data.census.gov, 2023.

References

[1] KuroSaru. QNX6 parser, 7 2018.

[2] Mozilla Foundation. Privacy nightmare on wheels: Every car brand reviewed by Mozilla, including Ford, Volkswagen, and Toyota, flunks privacy test, 2024.

[3] U.S. Census Bureau. Vius211a: All vehicles by registration state, vehicle type, and trailer configuration for the u.s. (excluding new hampshire) and states: 2021. Vehicle inventory and use survey, U.S. Census Bureau, 2021. ECNSVY Vehicle Inventory and Use Survey All Vehicles. Accessed on October 2, 2023. Data suppressed for quality reasons is blanked out.