

Kurzfassung zum Schlussbericht

zum Vorhaben

Konzeption und RT-kompatible Erweiterung von Zentralrechenplattformen und Embedded Compute Netzwerken für zukünftige hochautomatisierte Fahrzeuge (EMDRIVE)

FÖRDERKENNZEICHEN 16ME0455



FZI Forschungszentrum Informatik

Haid-und-Neu-Str. 10-14

76131 Karlsruhe

Autoren: Victor Pazmino Betancourt, Marius Kreutzer, Maximilian Kirschner

Telefon: +49 721 9654-190

E-Mail: pazmino@fzi.de

Die Systemarchitekturen im Automobil befinden sich in einem radikalen Wandel. Trends wie das autonome Fahren, Elektromobilität und weitreichende Vernetzung führen zu einer Abkehr von traditionellen Architekturen mit einer Vielzahl einzelner Steuergeräte (ECUs). Künftige Fahrzeug-Bordnetze zeichnen sich stattdessen durch eine stärkere Zentralisierung aus, bei der Hochleistungs-Car-Server und Domänen-Controller Applikationen bündeln und virtualisieren. Parallel dazu wird Rechenleistung dezentral in der Nähe von Sensoren und Aktuatoren integriert, um Daten vorzuverarbeiten und die Kommunikationsnetzwerke zu entlasten.

Diese Entwicklung stellt enorme Anforderungen an Energieeffizienz, funktionale Sicherheit (Safety) und Angriffssicherheit (Security). Insbesondere die Gewährleistung von höchsten Sicherheitsleveln wie ASIL-D erfordert extrem zuverlässige Systeme. Die forschungsleitende Hypothese des Projekts EMDRIVE war, dass zur wirtschaftlichen Realisierung solch komplexer Systeme eine dynamische Allokation der vorhandenen Rechenressourcen notwendig ist, um die Hardware-Redundanz im System zu minimieren.

Das Teilprojekt adressierte diese Herausforderung durch die Entwicklung von dynamischen Betriebsstrategien, die es ermöglichen, Software-Aufgaben zur Laufzeit flexibel auf die am besten geeignete Hardware zu verlagern und so die Effizienz, Resilienz und Sicherheit des Gesamtsystems zu steigern.

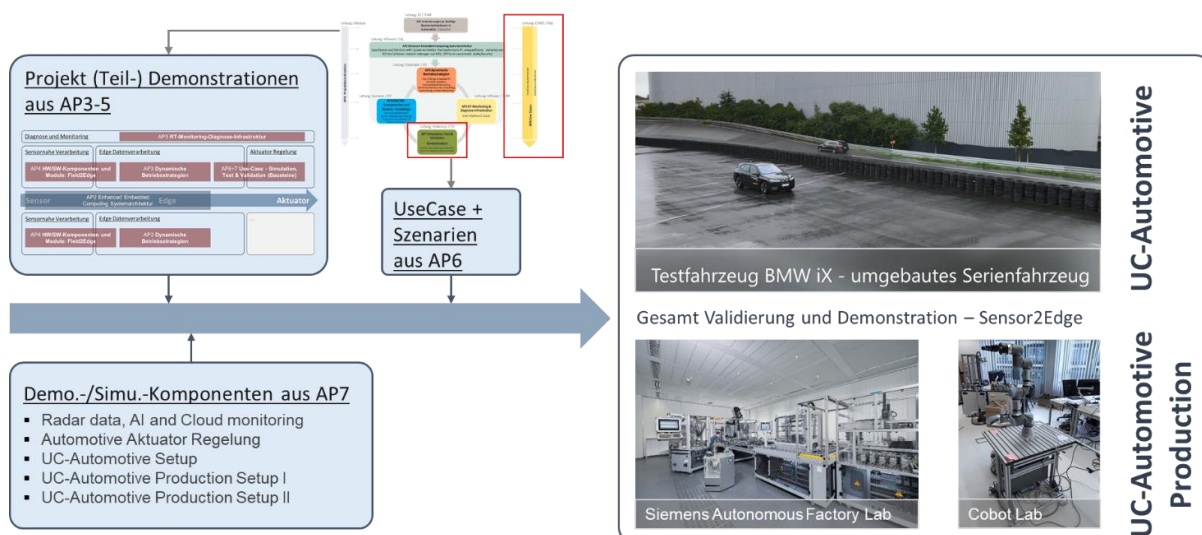


Abbildung 1: Gesamtkonzept zur Integration der Projektergebnisse

Die im Projekt entwickelten Konzepte wurden in einer modularen Demonstrationskette validiert. Diese integriert die technologischen Bausteine aus den Kernarbeitspaketen (AP3-5) und bettet sie in die Szenarien aus den Anwendungsfällen (AP6) ein. Dadurch konnte der Gesamterfolg des Projekts anhand von praxisnahen Demonstratoren nachgewiesen werden, die von der Simulation über Laboraufbauten bis zum Testfahrzeug reichten.

Ziele und Ergebnisse des Teilvorhabens

Die Arbeiten des FZI im Projekt EMDRIVE konzentrierten sich auf die Entwicklung und Umsetzung von dynamischen Betriebsstrategien für zukünftige, heterogene Fahrzeugarchitekturen. Die zentralen Ergebnisse lassen sich wie folgt zusammenfassen:

Entwicklung eines Frameworks zur Migration von Anwendungen mittels WebAssembly (WASM): Ein Forschungsschwerpunkt war die Entwicklung einer neuartigen Methode, um gekapselte Software-Anwendungen zur Laufzeit zwischen verschiedenen Recheneinheiten zu migrieren. Dies ermöglicht es, flexibel auf Lastspitzen zu reagieren, Energie zu sparen oder Anwendungen von einem fehlerhaften Knoten zu verschieben.

- Die Basistechnologie WebAssembly (WASM) sorgt für eine plattformunabhängige und sichere Ausführung in einer Sandbox, die Anwendungen strikt voneinander isoliert.
- Durch den Ansatz der Ahead-of-Time (AOT) Kompilierung wird eine performante Ausführung auf heterogenen CPU-Architekturen (z. B. von x86 auf ARM) ermöglicht.
- Ein mehrstufiger Migrationsprozess reduziert die Latenz im kritischen Pfad signifikant, indem zuerst der unveränderliche Code und erst im entscheidenden Moment der dynamische Zustand der Anwendung übertragen wird. Validierungsmessungen zeigten eine durchschnittliche Migrationszeit von nur ca. 51 ms zwischen heterogenen Systemen.

Konzeption einer generischen Schnittstelle zur dynamischen Beschleuniger-Anbindung: Moderne Systeme nutzen spezialisierte Hardware (z. B. FPGAs) für KI-Anwendungen. Um deren Auslastung zu optimieren, wurde eine service-orientierte Schnittstelle konzipiert, die es mehreren Anwendungen erlaubt, sich die verfügbaren Beschleuniger dynamisch zu teilen.

- Die Architektur basiert auf der Middleware ROS2 und besteht aus einem zentralen Hardware Manager, der Anfragen koordiniert, und spezifischen "Acceleration Runners", die als standardisierte Plugins für die jeweilige Hardware fungieren.
- Durch die Verwendung des Open Neural Network Exchange (ONNX) Formats wird eine hohe Kompatibilität von KI-Modellen über verschiedene Hardware-Plattformen hinweg sichergestellt.

Integration und Demonstration in realitätsnahen Anwendungsfällen: Die entwickelten Technologien wurden in einem Gesamtsystem integriert und validiert. Es wurde eine Ursache-Wirkungs-Kette demonstriert, bei der eine von der Diagnose-Infrastruktur (AP5) erkannte Leistungsdegradation als Auslöser für das dynamische Framework (AP3) dient, um eine kritische Anwendung auf eine andere Recheneinheit zu migrieren. Im Anwendungsfall der "Automotive Produktion" wurde die dynamische Migration von Analysealgorithmen zwischen Edge-Knoten und zentralen Servern zur vorausschauenden Wartung erfolgreich umgesetzt.

Die erzielten Ergebnisse bilden die Grundlage für drei Dissertationen, von denen eine im Rahmen des Projektes erfolgreich abgeschlossen wurde. Zusätzlich wurden im Rahmen des Vorhabens 5 studentische Abschlussarbeiten betreut. Die Ergebnisse wurden auch in Form von vier wissenschaftlichen Publikationen auf internationalen Konferenzen vorgestellt und werden in aktuellen Forschungsprojekten aufgegriffen und weiterentwickelt.