A Distributed Architecture for Runtime Fault Management in RISC-V based edge-AI SoCs

Ashwin Santhosh[1], Artur Jutman[1,2], Endri Kaja[3], Wolfgnag Ecker[3], and Maksim Jenihhin[1]

[1]Department of Computer Systems, Tallinn University of Technology, Tallinn, Estonia
[2]Testonica Lab OU¨ , Tallinn, Estonia
[3]Infineon Technologies AG, Munich, Germany

Abstract

Modern embedded processors are increasingly used in applications where reliability and safety are critical. At the same time, these systems are becoming more complex and are exposed to faults caused by manufacturing variations, aging, and transient effects. Traditional fault-tolerance techniques are often intrusive, static, or limited to isolated components, which reduces their suitability for adaptive, system-level reliability management. To address these challenges, this work introduces the concept of Doctor Core, a centralized decision and diagnosis engine intended to enable closed-loop reliability supervision in modern processors. Doctor Core continuously analyzes health information propagated from the underlying hardware, correlates fault symptoms across execution stages, and supports informed mitigation and recovery decisions at runtime. As a first implementation step toward this vision, we present NURSE (Non-Intrusive Unified Reliability Supervision Engine), a lightweight and distributed runtime monitoring framework integrated into the processor microarchitecture. NURSE performs stage-aware supervision by observing data and control flows within the pipeline, generating reliability metadata, and propagating health information across execution stages without modifying the functional behavior of the processor. As a proof of concept, the proposed approach is integrated into the open-source Ibex RISC-V processor, demonstrating how distributed monitoring can be embedded with minimal intrusiveness while enabling systematic runtime fault detection, localization, and health reporting to the higher-level decision layer. Together, Doctor Core and NURSE establish a scalable foundation for adaptive reliability management in future safety-critical and edge-AI computing systems.