UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

DOTTORATO DI RICERCA / PHD PROGRAM IN
INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING

Seminar announcement

Thursday 27 June 2024, Time: 10:00 - 11:00

Room T.1, Floor 0, Building 1, Via Claudio, 21 – NAPOLI

https://teams.microsoft.com/l/meetup-join/19%3ameeting_MDIyYzkxZGUtYzZhZS00NmY2LTg3ODgtMDNmN2Q5MWQ3MGY5%40thread.v2/0?context=%7b%22Tid%22%3a%222fcfe26a-bb62-46b0-b1e3-28f9da0c45fd%22%2c%22Oid%22%3a%227630d56e-adf1-4b76-87a9-d5997c46f95f%22%7d

Dr-Ing. Gautam Gala

Post Doctoral Researcher

Technical University of Kaiserslautern, Germany

Resource management and orchestration for mixed-criticality cloud/distributed systems

Abstract: Mixed-criticality distributed systems implement several applications of different criticality levels on a single shared Commercial-Off-The-Shelf (COTS) multicore processor to help reduce Size, Weight, and Power (SWaP), required wiring, and associated costs. A new trend is emerging, where real-time safety-critical systems are offloaded onto cloud/edge platforms to leverage the benefits of cloud computing. Similar to distributed systems, multiple applications run on each cloud node and share the underlying node resources.

It is challenging to achieve the required isolation in such distributed or cloud/edge nodes, mainly when using multicore nodes, due to contention in the shared resources such as CPU, shared bus, memory (controller), and network. These shared resources can cause unpredictable delays, leading to deadline misses in real-time applications. Moreover, the boundaries for fault isolation and error contamination are not sharply defined. There are also conflicting goals for safety-critical and non-critical/best-effort applications that exacerbate the preexisting problem. The pessimistic Worst-case Execution Time (WCET) estimations of safety-critical applications under-utilize the resources significantly in the average case. Simultaneously, the non-critical/best-effort applications require efficient resource utilization to provide the best possible Quality of Service (QoS). Existing approaches built into hypervisors or OSes assume constant availability and amount of CPU processing power to provide guaranteed progress of tasks. Such pre-planned assumptions may be feasible on single nodes; however, they become less meaningful in distributed systems or cloud/edge environments with multiple nodes, each having multiple shared resources. If the applications, availability of resources, or system configurations change, obtaining assumptions about resources becomes complicated. Moreover, it is challenging to meet end-to-end constraints by considering each resource or node individually.

We present our resource management techniques, which help ensure we meet deadlines of real-time applications, guarantee isolation and predictable shared resource access delays, and contributes towards providing the best possible QoS to non-critical/best-effort tasks. Our global resource management and orchestration framework helps maintain a system/cloud-wide view of resources and applications. It coordinates and adapts system/cloud-wide resource allocations even when resource availability and requirements change dynamically.

Lecturer short bio: *Gautam Gala completed his Ph.D. in real-time systems (EIT) at TU Kaiserslautern-Landau (RPTU), Germany in 2021 and continued as a post-doctoral researcher. He has collaborated with many industries and European universities on several scientific publications. He worked on industrial and EU projects such as DREAMS and SECREDAS. He is co-organizing the 3rd workshop on RT-cloud systems in 2024 (co-hosted with ECRTS). His research is focused on resource management and orchestration for safe and predictable cloud-based or distributed real-time (RT) / mixed-criticality systems, using cloud technology (such as containers and Kubernetes) for RT systems, and low-level monitoring on multicore platforms. His technical expertise spans several programming languages and technologies, including Linux kernel programming, cloud orchestrators (e.g., Kubernetes), virtualization (e.g., KVM, Xen, XtratuM, PikeOS, Docker), RTOSes (e.g., FreeRTOS).*

For information: Prof. Marcello Cinque (DIETI, UniNA) – marcello.cinque@unina.it