UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

itee PhD
information technology
electrical engineering

DIETI UNINA

Giorgio Farina

Improving Isolation in Real-Time Cloud Platforms

Tutor: Prof. Marcello Cinque

Cycle: XXXVII

Year: 2023/2024

itee PhD
information technology
electrical engineering

Candidate’s information

- **MSc degree:** Computer Engineering
- **DIETI Research group/laboratory:** DESSERT LAB
- **PhD start date:** 2021 – **end date:** 2024
- **Scholarship type:** CINI
- **Period abroad:** Purdue University (8 months), West Lafayette, Indiana, USA

Summary of study activities

- **Attended Courses:**
    - Software Security
    - Real-Time Industrial Systems
    - Virtualization technologies and their applications
    - Statistical data analysis
    - IoT Data Analysis
- **Attended PhD Schools:**
    - “Verification and Validation of Automated Systems’ safety and Security”

Research area(s)

- My research area during these three years mainly included the testing, methods and design of dependable, safety and resilient solutions in cyber physical systems
- I focused on the problem of isolating the execution of different tasks in modern computer architectures
  - **1° year:** Memory access isolation
  - **2° year:** Software isolation in Cloud
  - **3° year:** Confidential Execution Environments in Android

Thesis

Research results

- **Evaluation of the regulation enforced by the hardware controller Intel MBA**
- **Memory queue occupancy as metric to detect the memory access interference**
- **A record and replay framework for hypervisor intervention in cloud**
- A programmer-friendly confidential execution evironment in Android
- A tool for testing CPU virtualization fault-tolerance mechanisms in hypervisors with an empirical evaluation on three different hypervisors

Thesis

Research products

[P1]	Giorgio Farina, Gautam Gala, Marcello Cinque, Gerhard Fohler Enabling memory access isolation in real-time cloud systems using Intel’s detection/regulation capabilities, International Journal of Systems Architecture Volume 137, April 2023, 102848, DOI: 10.1016/j.sysarc.2023.102848
[P2]	Giorgio Farina, Gautam Gala, Marcello Cinque, Gerhard Fohler, Assessing Intel’s Memory Bandwidth Allocation for resource limitation in real-time systems, IEEE 25th International Symposium On Real-Time Distributed Computing (ISORC), Västerås, Sweden, 17-18 May 2022, IEEE, DOI: 10.1109/ISORC52572.2022.9812757
[P3]	Carmine Cesarano; Marcello Cinque; Domenico Cotroneo; Luigi De Simone; Giorgio Farina IRIS: a Record and Replay Framework to Enable Hardware-assisted Virtualization Fuzzing, 2023 53rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), Porto, Portugal, 27-30 June 2023, IEEE, DOI: 10.1109/DSN58367.2023.00045

Research products

[P4]	Marco Barletta, Marcello Cinque, Luigi De Simone, Raffaele Della Corte, Giorgio Farina, Daniele Ottaviano RunPHI: Enabling Mixed-criticality Containers via Partitioning Hypervisors in Industry 4.0, 2022 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), Charlotte, NC, USA, p. 134-135, 26 December 2022, IEEE, DOI: 10.1109/ISSREW55968.2022.00058
[P5]	Marco Barletta, Marcello Cinque, Luigi De Simone, Raffaele Della Corte, Giorgio Farina, Daniele Ottaviano Partitioned Containers: Towards Safe Clouds for Industrial Applications, 2023 53rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks-Supplemental Volume (DSN-S), Porto, Portugal, p. 84-88, June 2023, IEEE, DOI: 10.1109/DSN-S58398.2023.00029
[P6]	Marcello Cinque, Raffaele Della Corte, Giorgio Farina, Stefano Rosiello An unsupervised approach to discover filtering rules from diagnostic logs, 2022 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), Charlotte, NC, USA, p. 1-6, 26 December 2022, IEEE, DOI: 10.1109/ISSREW55968.2022.00030
[P7]	Marcello Cinque, Raffaele Della Corte, Giorgio Farina, Stefano Rosiello AID4TRAIN: Artificial Intelligence-Based Diagnostics for TRAins and INdustry 4.0, European Dependable Computing Conference 2022 Workshops, Zaragoza, Spain, vol 1656, p. 1-6, 05 September 2022, Springer, Cham, DOI: 10.1007/978-3-031-16245-9_7

PhD thesis overview

- **PhD Thesis vision**

  Porting critical apps in cloud

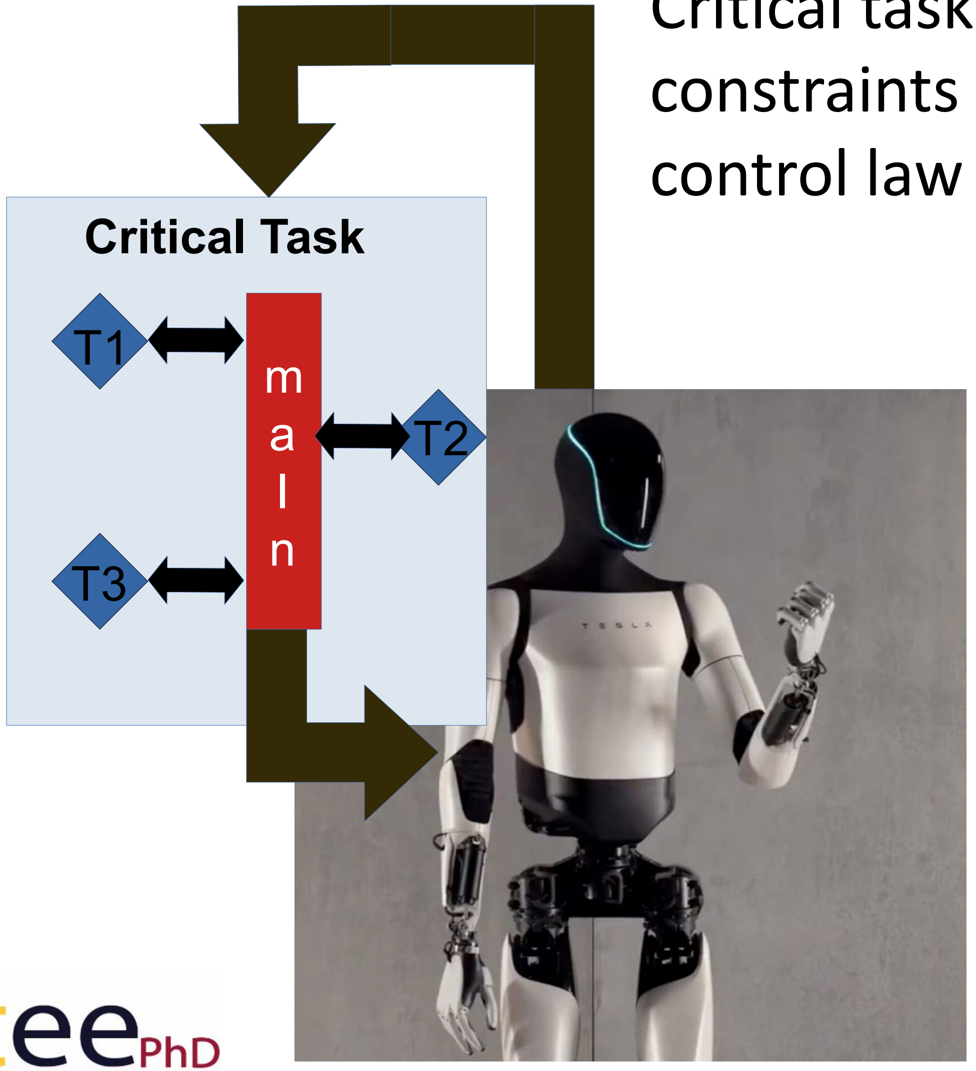
  To improve utilization and reduce running costs

- **PhD Thesis Objective**

  Overcoming some of the threats in cloud platforms that limit the RT-Cloud vision

- **Methodology**
  - Black box evaluation of hardware controllers
    - We studied the traffic patterns traversing the hardware controller to design the synthetic workloads
  - Queue occupancy as detector of memory access interference
    - We started from an intuition and we provide evidence empirically
  - IRIS
    - We applied a well-known technique (Record and replay) to a specific problem
    - We evaluated the accuracy and efficiency comparing the replied execution with the recorded execution

Context



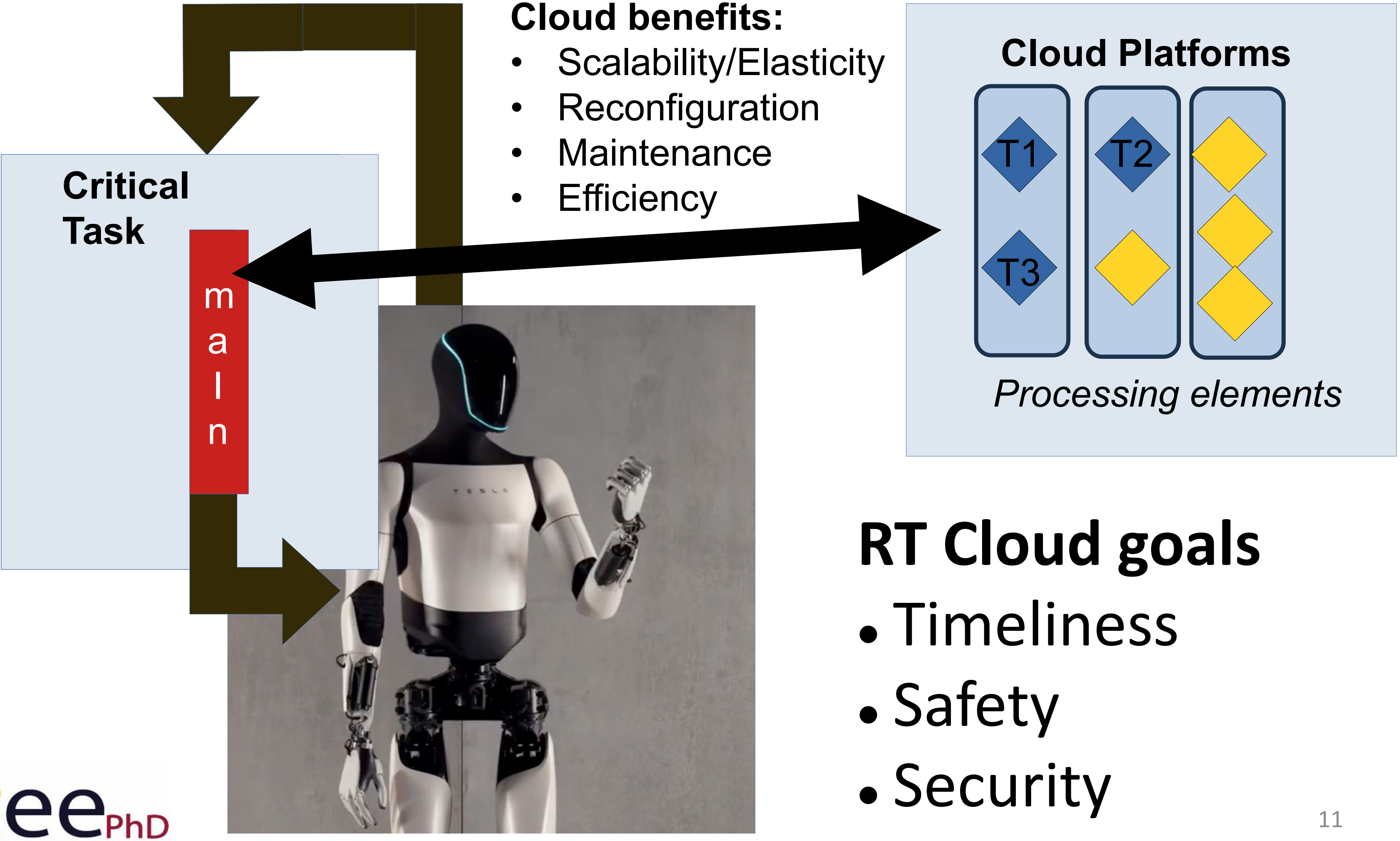
Critical tasks may include time-constraints (real-time) such as the control law of a robot

The failures of critical tasks (e.g., deadline miss) have high impact on the system and on the environment

In the last decade, **the software complexity and the costs of critical tasks increased...**

Real-Time Cloud

Vision: “porting critical apps in cloud”



Cloud benefits:

- Scalability/Elasticity
- Reconfiguration
- Maintenance
- Efficiency

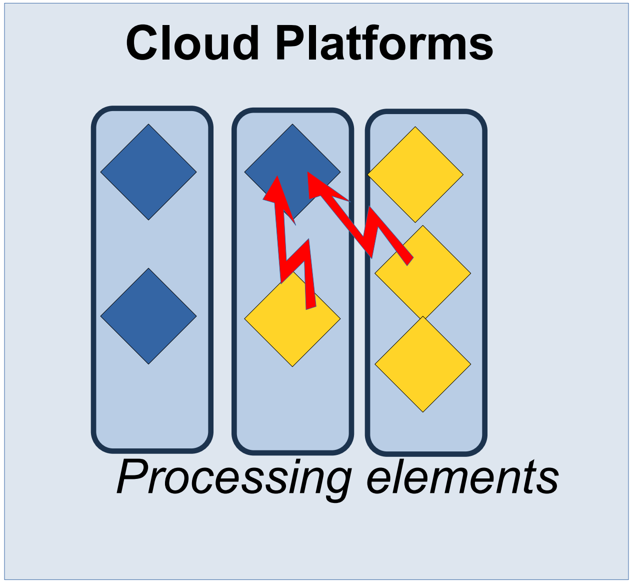
RT Cloud goals

- Timeliness
- Safety
- Security

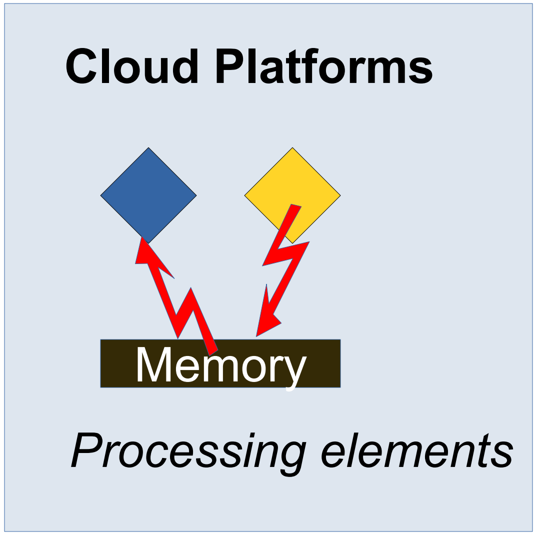
Real-Time Cloud

Threats to guarantees:

- shared hardware resource interference
  - **Memory access isolation**
- shared software layer interference
  - **Hypervisor robustness**



Memory Access Isolation

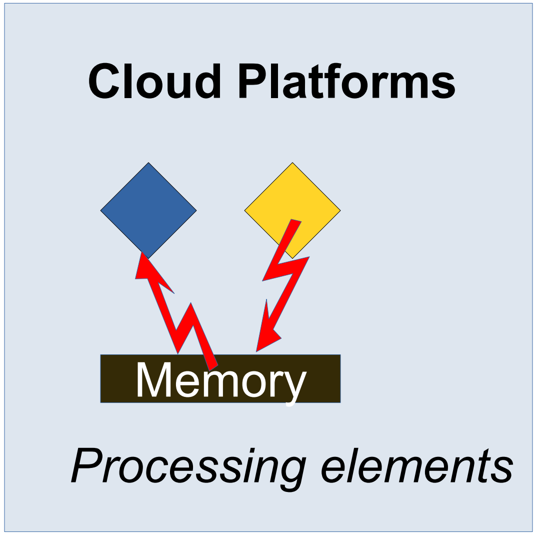


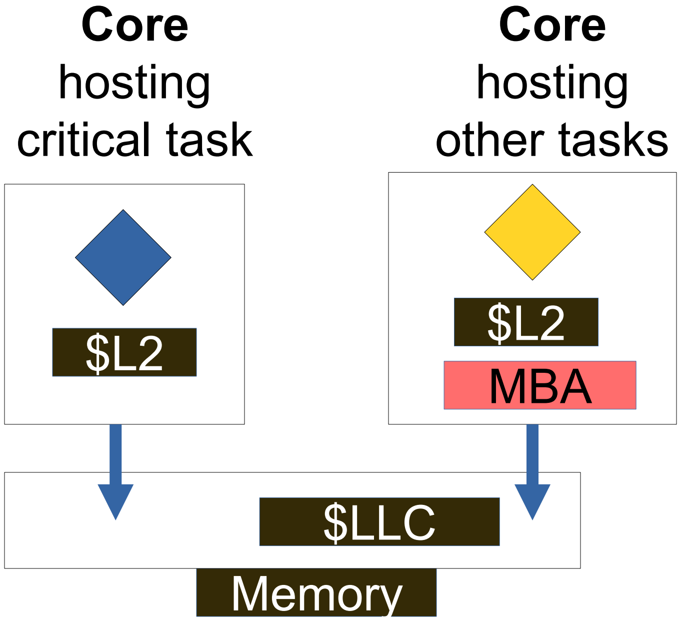
Problem: *"non-critical tasks of other customers may interfere with the critical tasks by contending the memory access"*

Memory Access Isolation

Intel Memory Bandwidth Allocation:

it is a ***black box hardware controller*** to delay requests going to the high-speed interconnect of a Intel Xeon Scalable Servers





What is the MBA indirect limitation on the memory bandwidth of the non-critical tasks?

Memory Access Isolation

Finding 1: Our workload “exclusive DRAM Bomb” is capable of **bypassing the regulation shown by generic state-of-the-art workloads by over 50%,** warning the community about the risks of adopting generic workloads for specialized controllers **[1]**

Our workload: “exclusive DRAM Bomb”

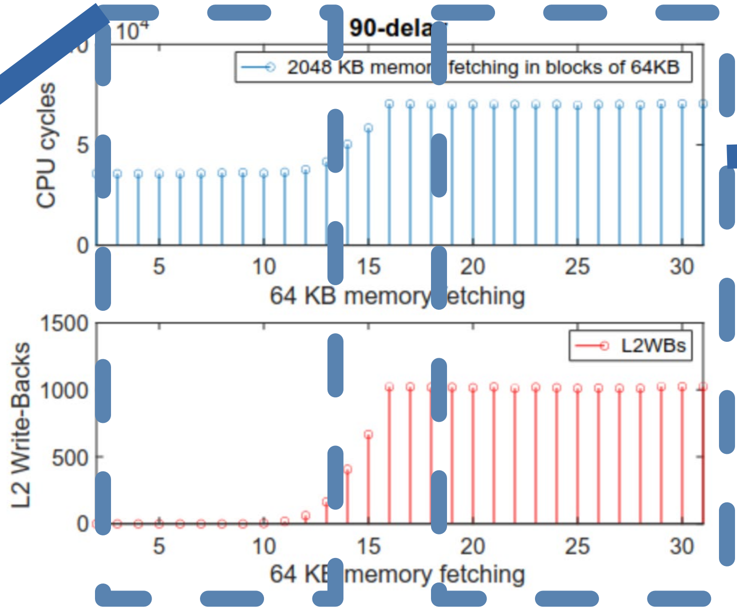


Figure: *2048 KB memory fetching in several blocks of 64KB when MBA 90-delay is enabled*

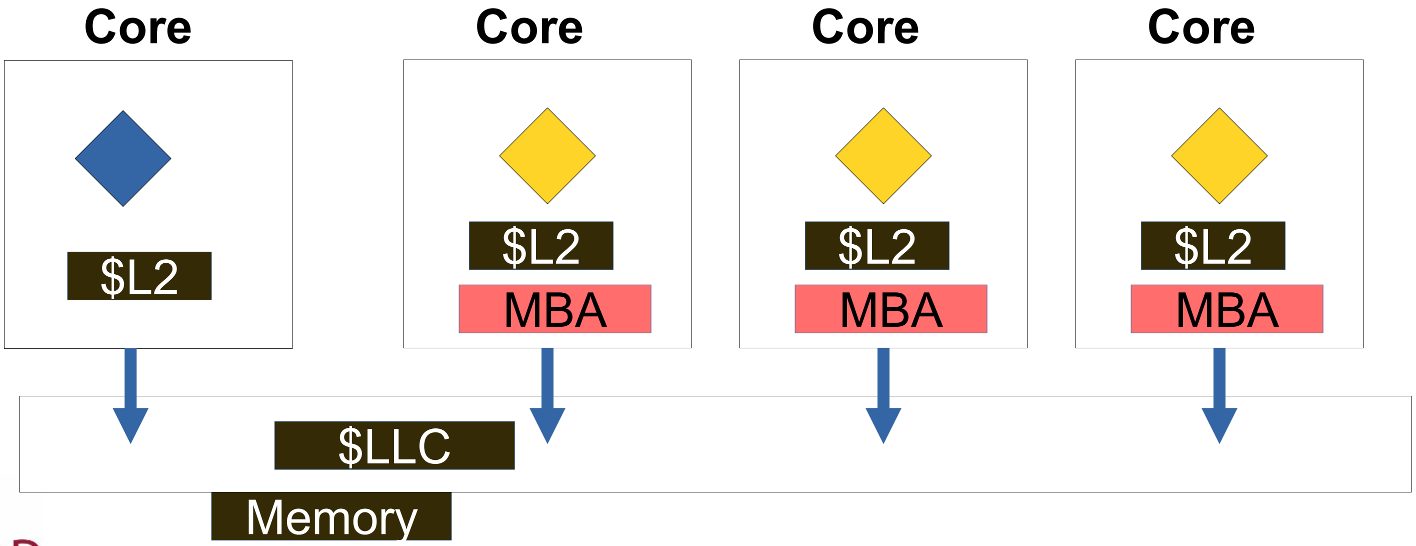
State-of-art: Generic memory Workloads (***e.g., DRAM bomb***) adopted by the related works

[1] Giorgio Farina, Gautam Gala, Marcello Cinque, Gerhard Fohler, “Assessing Intel’s Memory Bandwidth Allocation for resource limitation in real-time systems”, 2022 IEEE 25th International Symposium On Real-Time Distributed Computing (ISORC),

Memory Access Isolation

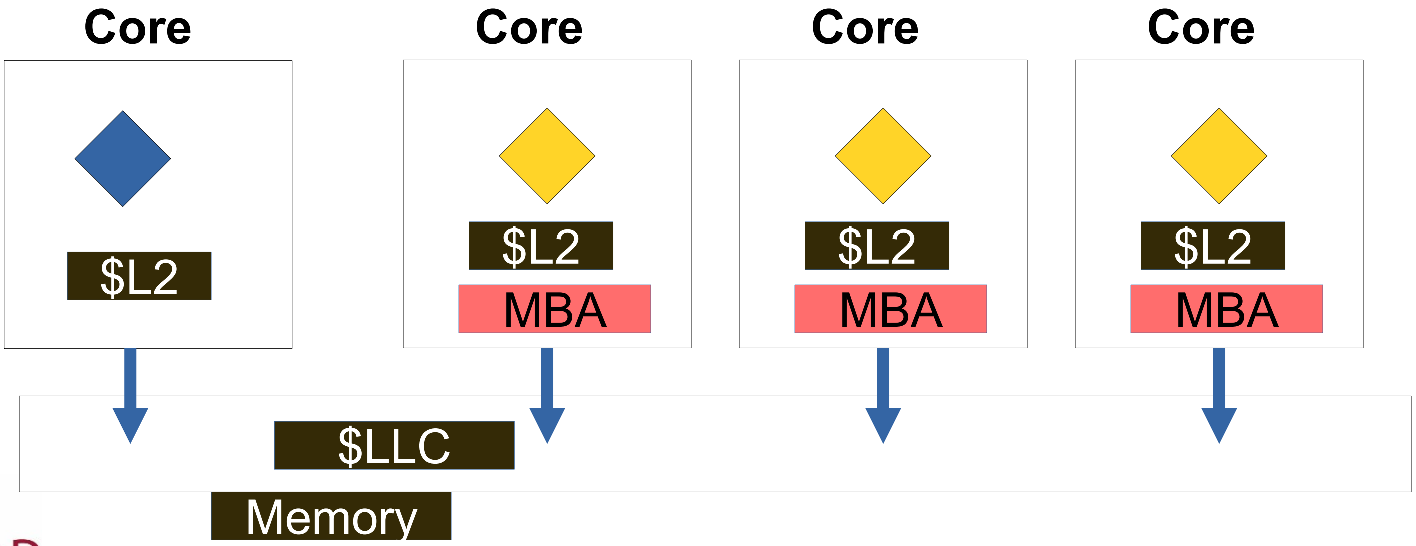
Problem: *“MBA limits non-critical tasks regardless of the current interference, affecting the memory bandwidth utilization”*

How can we detect memory access interference?



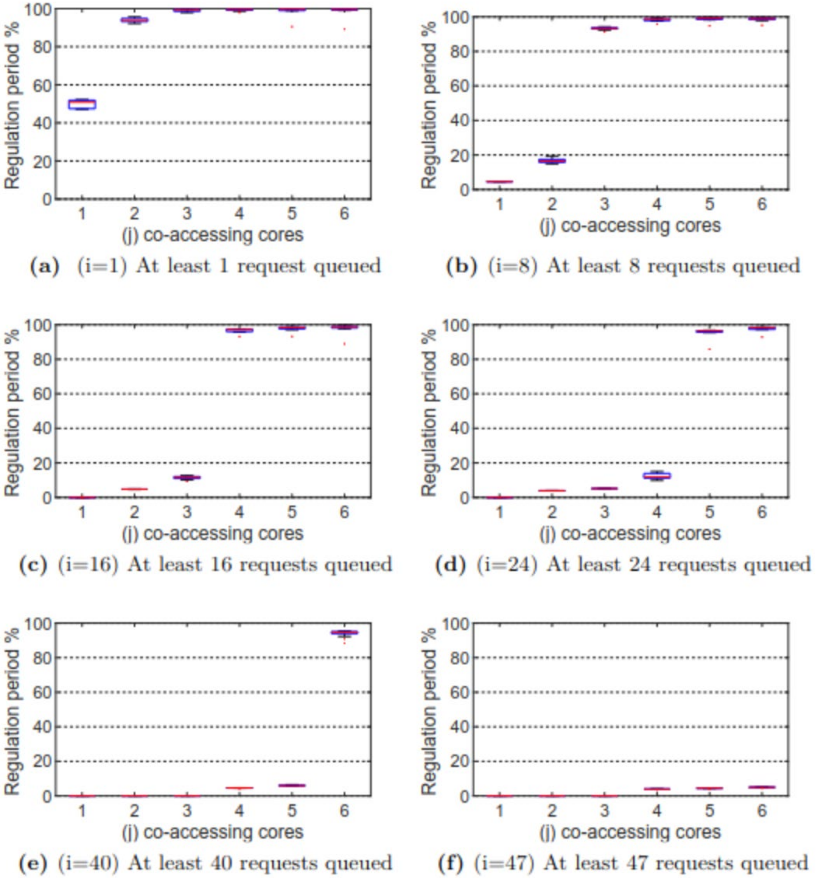
Memory Access Isolation

Key intuition: *"By monitoring Memory Queue Occupancy (MQO), we can detect the degree of interference, i.e., the number of cores co-accessing to the memory"*



Memory Access Isolation

***Finding 2:** "Detecting more than two cores requires a detection time corresponding to one-fifth of the regulation period while detecting more than three, four, and five cores requires one-sixth **[2]**"*



(a) (i=1) At least 1 request queued

(b) (i=8) At least 8 requests queued

(c) (i=16) At least 16 requests queued

(d) (i=24) At least 24 requests queued

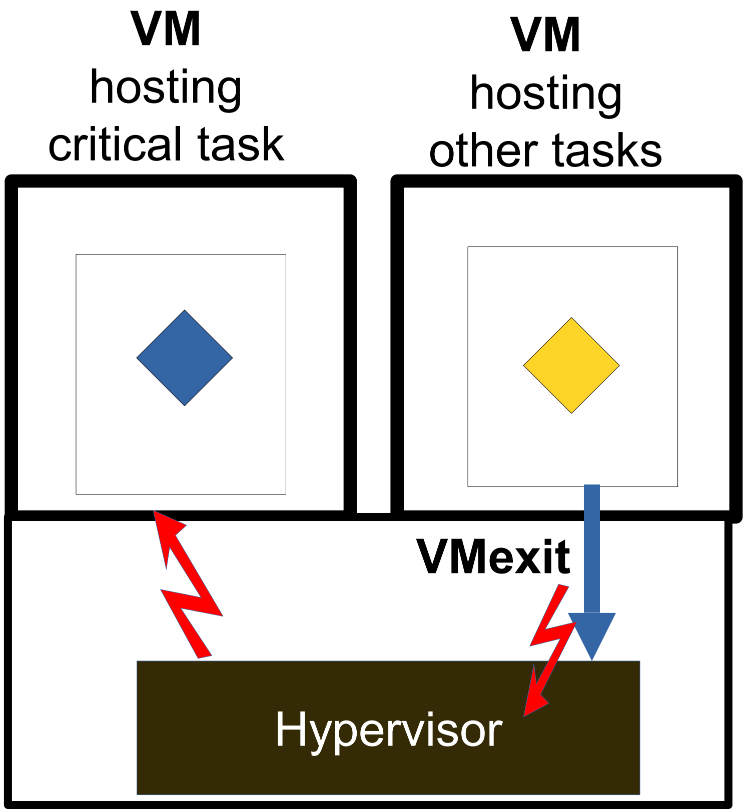
(e) (i=40) At least 40 requests queued

(f) (i=47) At least 47 requests queued

***Figure:** Portion of the regulation period in which the Read Pending Queue occupancy is at least **i** when **j** cores are co-accessing the memory*

[2] Giorgio Farina, Gautam Gala, Marcello Cinque, Gerhard Fohler,
"Enabling memory access isolation in real-time cloud systems using Intel's detection/regulation capabilities",
2023, Elsevier, International Journal of Systems Architecture (JSA)

Hypervisor robustness

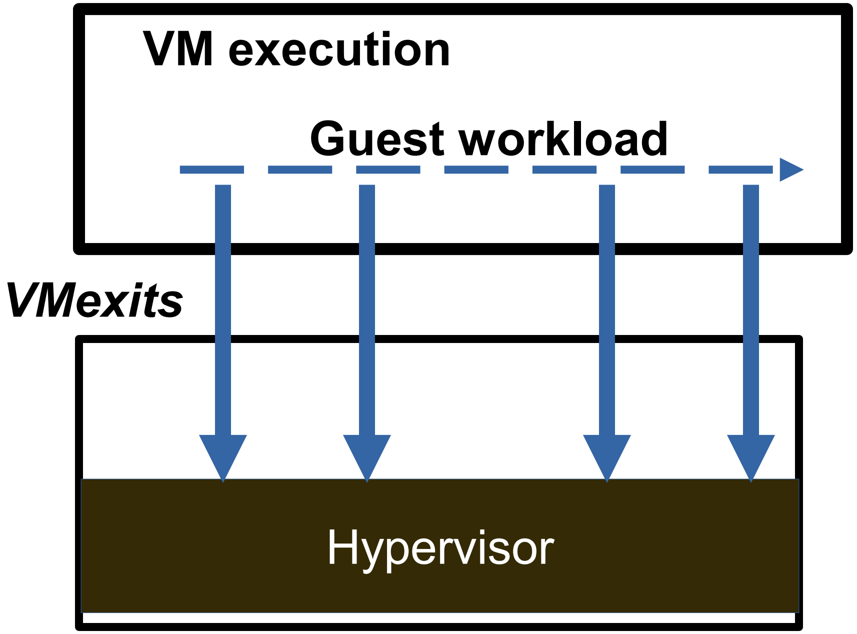


- **Context**
- The isolation properties of hardware virtualization are a key factor for porting critical applications in cloud
- Critical apps and non-critical apps can run into two distinct Virtual Machines (VM)
- Distinct virtual machines share a more privileged layer called *hypervisor*

- **Problem Statement**
- *A malicious VM can target functional or software bugs in the hypervisor in order to affect the execution of another VM*

How can we test the hypervisor intervention?

Secure CPU Virtualization

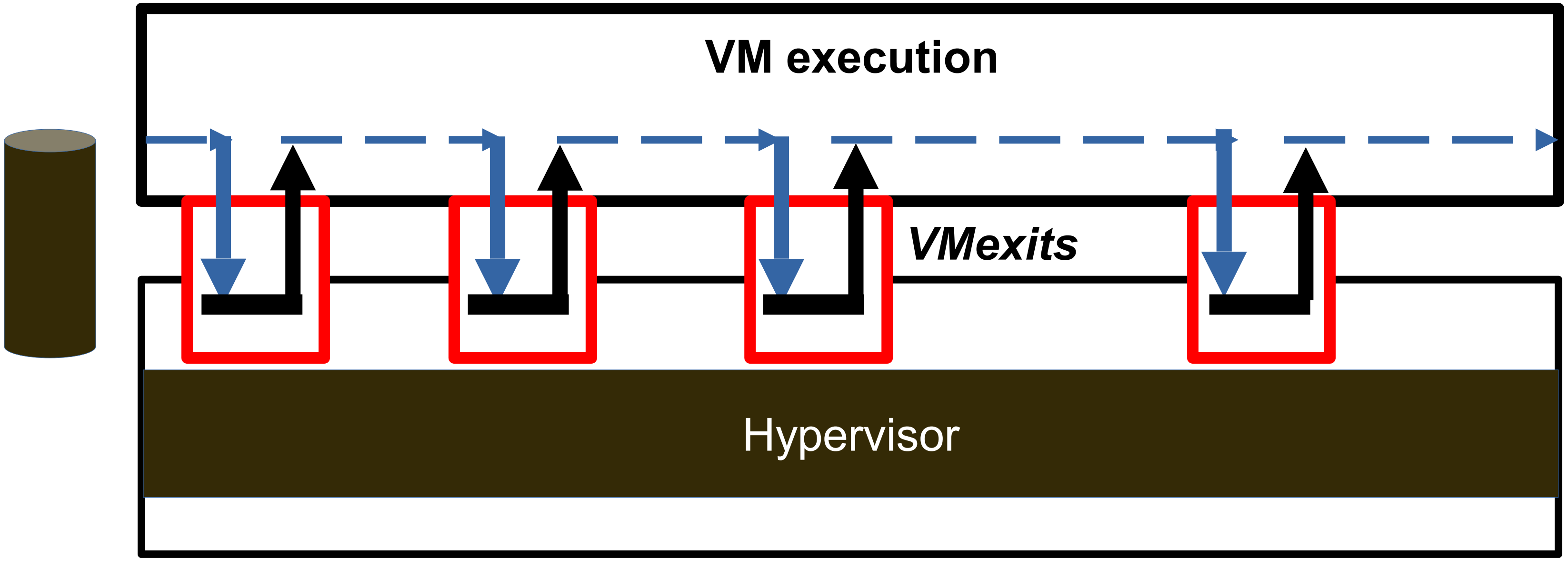


- **Challenges**
- Specific and configurable VM conditions trigger the hypervisor intervention
- Reproducing these VM conditions requires a deep knowledge of the underlying hardware (e.g., CPU specification)

Our key idea:
”instead of manually reproducing these complex guest workloads, we can replicate existing behaviors observed during the nominal execution of guest workloads”
For instance, we may record the hypervisor interventions during the boot of the guest OS.

Secure CPU Virtualization

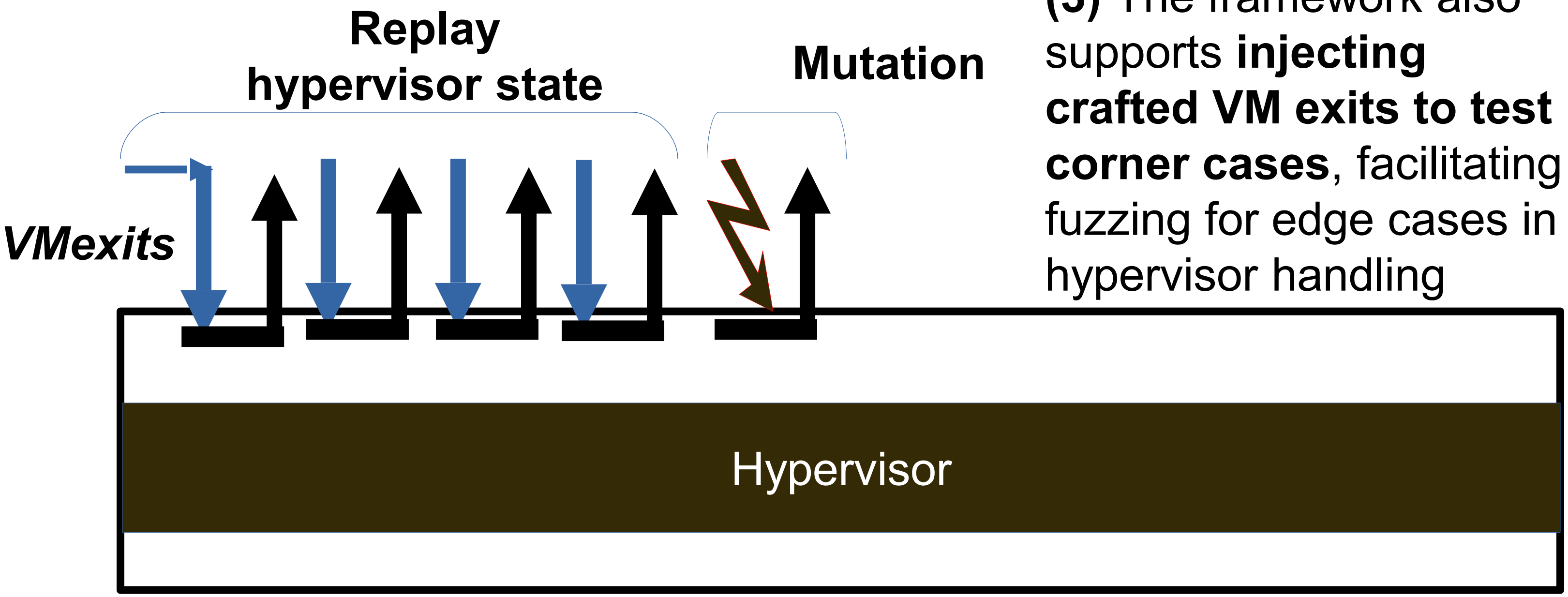
We propose **a record-and-replay** method



(1) We collect (learning) key events during **hypervisor intervention** by executing a guest workload.

Secure CPU Virtualization

We propose **a record-and-replay** method



(3) The framework also supports **injecting crafted VM exits to test corner cases**, facilitating fuzzing for edge cases in hypervisor handling

- **(2)** Our **replay** mechanism can submit the **recorded hypervisor behavior as a series of VM exits,** eliminating the need to re-execute guest workloads

Secure CPU Virtualization

We publicly released IRIS, a proof-of-concept implementation in Xen [3]

Finding 3:

- In our settings, compared to real guest execution, IRIS can reach the same valid hypervisor states replicating a series of VM exits,
  - With a fitting of code coverage ranging between 92,2% and 100%
  - With a time improvement from 42,5 to 99,6%

[3] Carmine Cesarano; Marcello Cinque; Domenico Cotroneo; Luigi De Simone; Giorgio Farina *"IRIS: a Record and Replay Framework to Enable Hardware-assisted Virtualization Fuzzing"*, 2023 53rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)