Conectando ideias, antecipando o futuro: inovação colaborativa para as redes 5G e 6G.

II WORKSHOP

Internacional do xGMobile

Realização:













Leveraging 6G Device-Edge-Cloud Orchestration of Networked Extended Reality

Sand Luz Correa

Instituto de Informática Universidade Federal de Goiás (UFG)















Agenda

- XR concept
- XR in 5G and beyond
- XR in 6G networks
- XR challenges
 - demand
 - scalability
 - orchestration







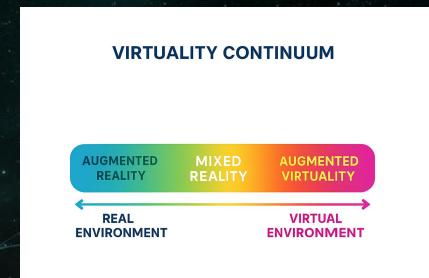


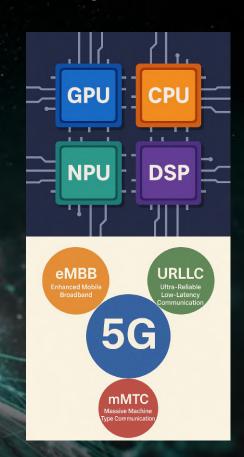






Extended Reality Concept















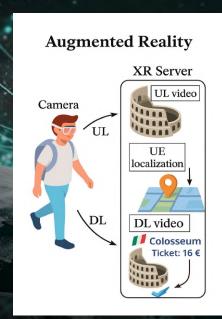


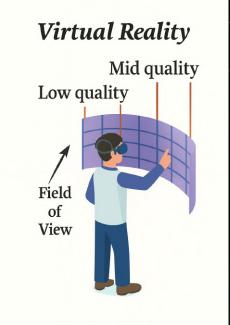


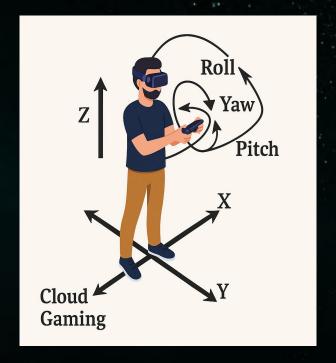


XR in 5G and Beyond

3GPP Use Cases



















XR in 5G and Beyond - Requirements

Traffic stream	DL/UL	Use Case	Packet rate	Av. data rate	Delay
Video	DL	AR	60 fps, [30,90,120] fps	30 - 60 Mb/s	10 ms
		VR			
		CG		8 - 45 Mb/s	15 ms
	UL	AR	60 fps	10 Mb/s, [20 Mb/s]	30 ms
			250 fps	0.2 Mb/s	10 ms
Motion/Control		VR			
		CG			
Audio + Data	DL+UL	AR/VR/CG in DL, AR in UL	100 fps	0.756 - 1.12 Mb/s	30 ms

ricauzação







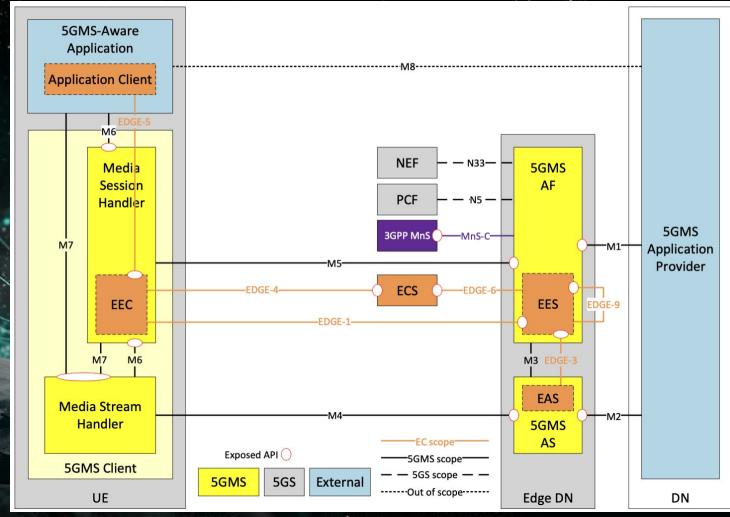






5G and Beyond Enabled XR Services

5G Media Streaming combined with Edge Computing













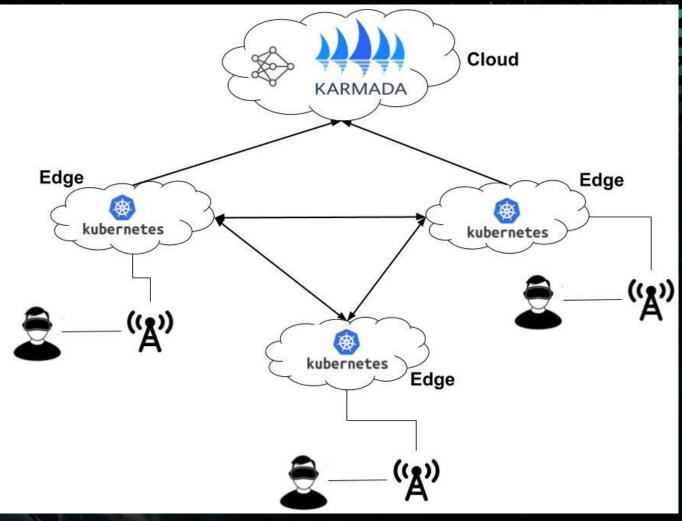




XR in 5G and Beyond - Orchestration

Container-based orchestration

Edge-Cloud Continuum











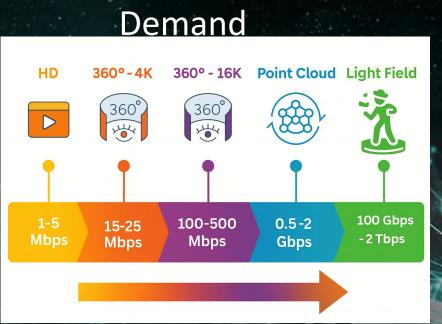








XR in 6G networks



Scalability

Portability











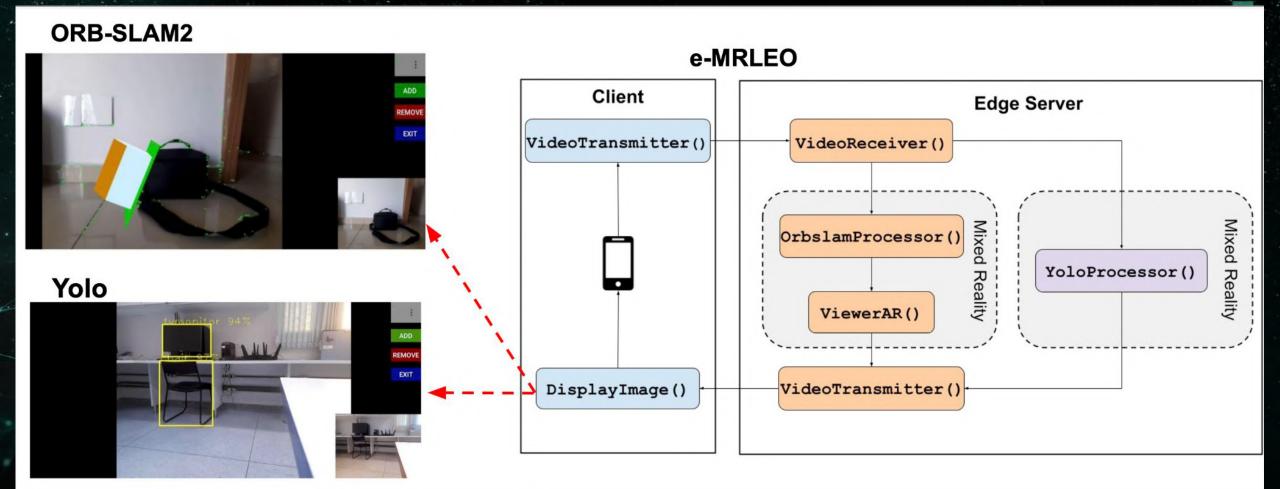






Understanding the Demand













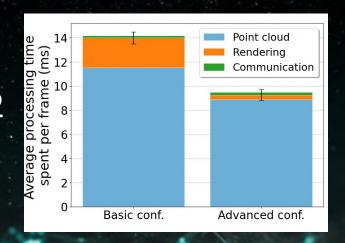


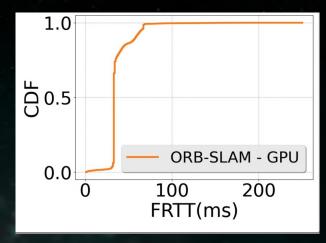


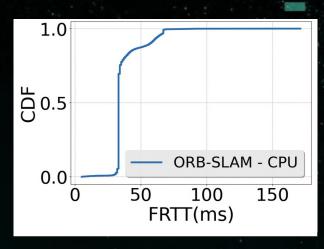


Understanding the Demand

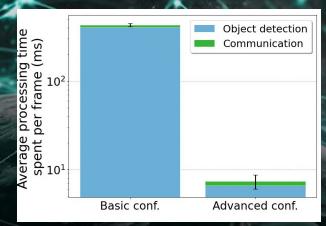
ORB-SLAM2







Yolo



1.0 — YOLO - CPU — YOLO - GPU 0.0 0 1000 FRTT (ms)

Realização:









GOVERNO MINISTÉRIO DA DE MINAS CIÊNCIA, TECNOLOGIA TREM PROSPERA. E INOVAÇÃO



Understanding the Demand

Task	Hardware	Activity	Model
		Task arrival	Periodically, every 33 ms
ORB-SLAM2	CPU	Computation	For 43%: $\mathcal{N}(\mu = 494 \text{MI}, \sigma = 147,97 \text{MI})$ For 53%: $\mathcal{N}(\mu = 404 \text{MI}, \sigma = 110,99 \text{MI})$
		Communication	39 kB per frame
		Memory used by edge server only	408 MB
		Additional memory for each user device	288 MB
		Task arrival	Periodically, every 33 ms
YOLO	GPU	Computation	For 82%: $\mathcal{GN}(\mu = 91 \text{MI}, \alpha = 0.40 \text{MI}, \beta = 0.80)$ For 18%: $\mathcal{GN}(\mu = 89 \text{MI}, \alpha = 2.05 \text{MI}, \beta = 1.25)$
		Communication	39 kB per frame
		Memory used by edge server only	208 MB
		Additional memory (RAM and VRAM) for each user device	416 MB for each memory















Evaluating Scalability

- Simulation is the main method of evaluating scalability
- Natively, none of the most popular Edge Computing simulation tools (EdgeCloudSim, PureEdgeSim, EdgeSimPy, iFogSim) includes an explicit GPU model
 - Number of SMs
 - GPU frequency
 - VRAM
 - o vGPU
 - Power model















Evaluating Scalability

CloudSim

EdgeCloudSim

GPUCloudSim

GPUEdgeCloudSim

Realização:



Inate







MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO

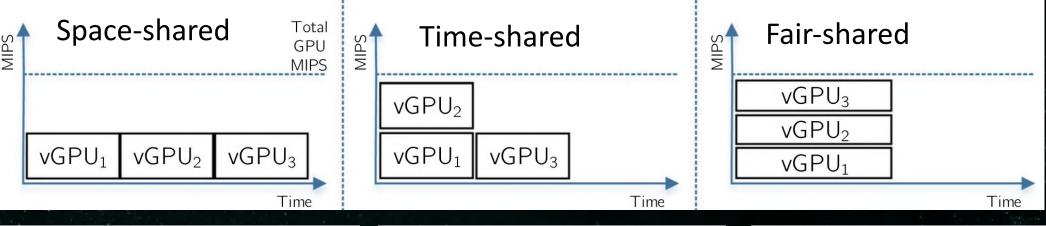
Application Application **CPU** CPU Task Task **GPU GPU** Task Task vCPU vGPU vRAM

VM

OS



GPU Provision



GPU
Task
Allocation

Leftover

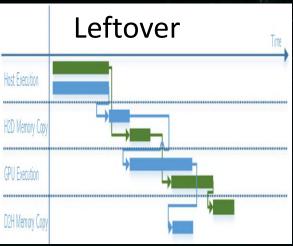
Host Execution

H2D Memory Copy

GPU Execution

D2H Memory Copy

D2H Memory Copy



VM Placement

First-Fit

Realização:









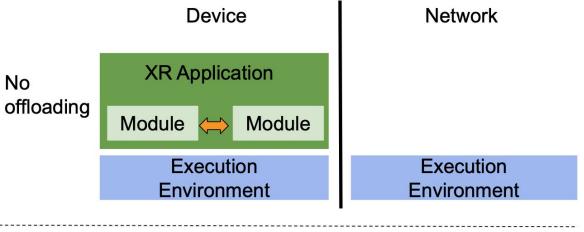


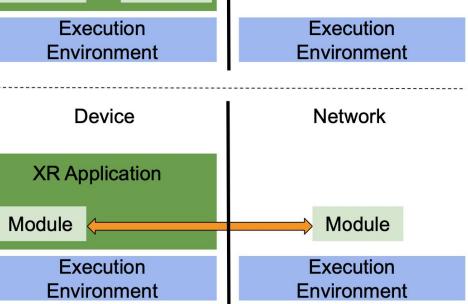


Design, implement and evaluate an offload framework that expand **XR applications** functionalities dynamically from devices to wireless connected compute

Addressing Portability







Offloading benefits

- reduce computational response time
- balancing compute and energy tradeoff
- reduce device heat

Design goals

- dynamic deployment and scheduling
- easy to consume programmatically
- lightweight
- platform-agnostic





offloading

into the

edge



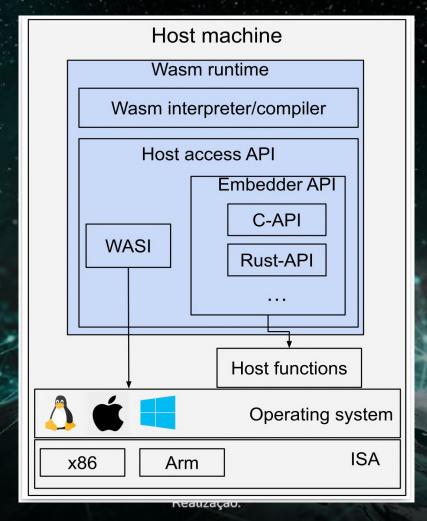


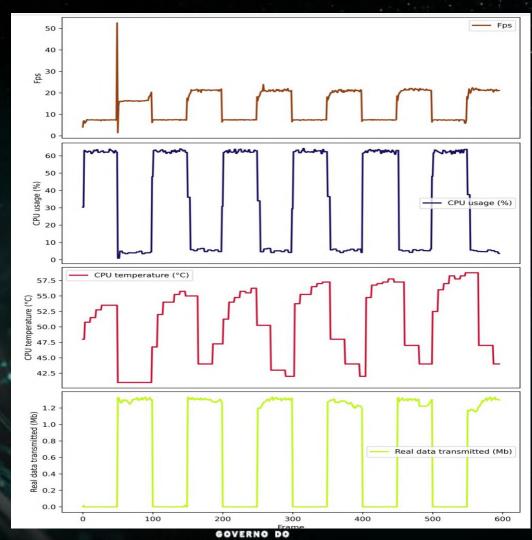






Addressing Portability

















II WORKSHOP

Internacional do xGMobile

sandluz@ufg.br

Realização:











