



*Heterogeneous Secure Multi-level Remote Acceleration
Service for Low-Power Integrated Systems and Developed*

*Cloud Forward 2016
19 October, Madrid*



Outline

- ▶ RAPID Focus
- ▶ Overview of the current situation
- ▶ Proposed solution
- ▶ Challenges and expectations
- ▶ Conclusions

Rapid Focus

- ▶ Novel heterogeneous CPU–GPU multi–level cloud acceleration
 - ▶ focusing on applications running on embedded systems found on low–power devices.
- Runtime system performs energy and performance estimations
 - to automatically select local CPU–based and GPU–based tasks
 - that should be seamlessly executed on more powerful remote devices or cloud infrastructures.
 - Secure unified model
 - where almost any device or infrastructure can operate as an accelerated entity and/or as an accelerator
 - serving other less powerful devices in a secure way



Overview of the current situation

Low-power devices and embedded systems cannot always cope with the increased demand for processing power, memory and storage required by modern applications

Most applications are only executed on high-end servers

Loading...

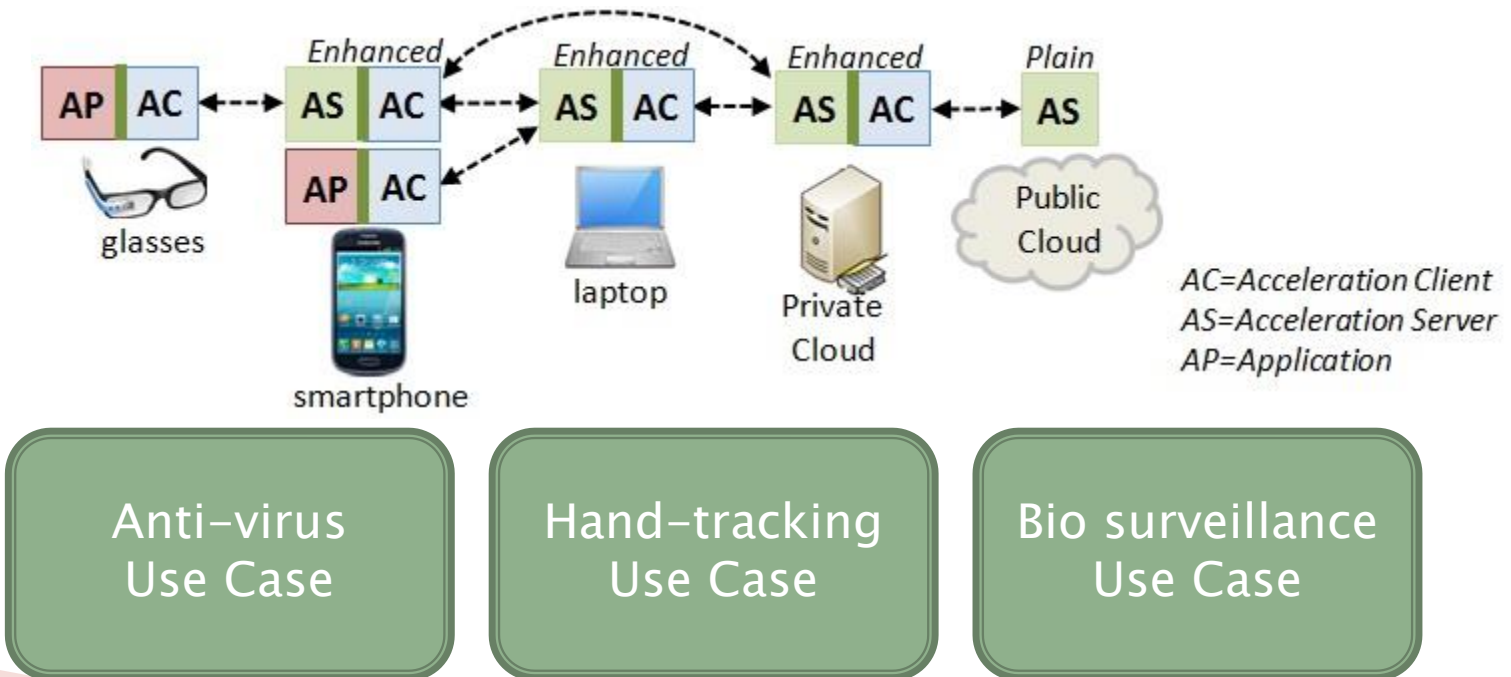


Users demand more features and services from mobile application developers

Offload heavy operations to more resource-full machines, such as private or public cloud
Mobile Cloud Computing.

Overview of the current situation

RAPID provides a full architecture of a heterogeneous offloading framework for making the system highly automatic and transparent to the developers and final users

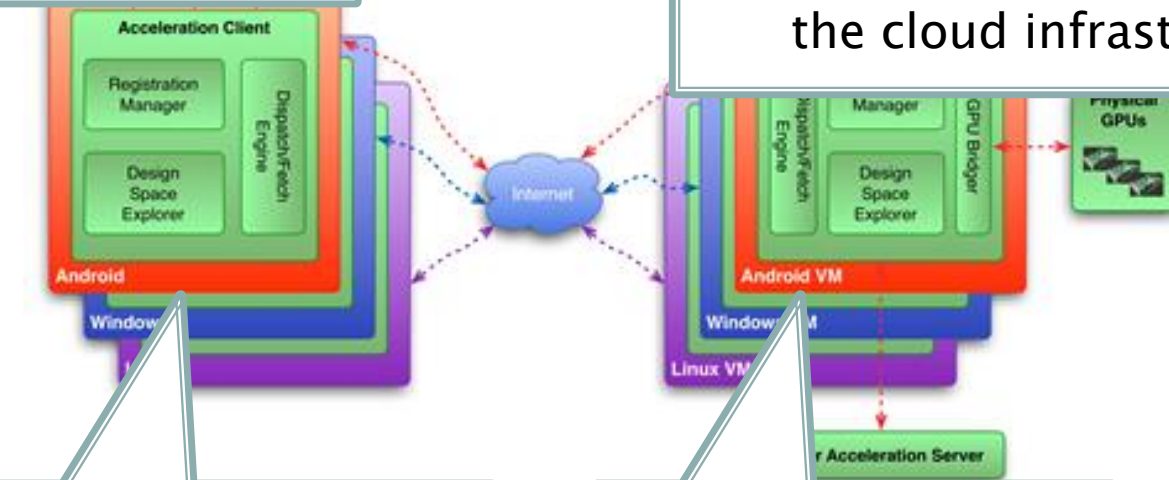


Proposed solution

VMM component responsible for managing and monitoring the resources of a physical machine

SLA Manager: QoS for Application Off-loading requirements

Directory Server centralizes the knowledge of computational resources and active entities in the cloud infrastructure



Acceleration Client handles the task of offloading process on the client side.

Acceleration Server is the component that runs inside a VM that handles the execution of the offloaded tasks.

Challenges and expectations

- Enable Mobile Cloud Computing Paradigm improving task offloading
- Quality-based offloading to remote hosts
- Increase computation capabilities
- Negotiation and management of SLAs
- Improve allocation mechanisms and resources management



RAPID proposes the first mechanism for accessing remote GPUs

Conclusions

RAPID proposes a new heterogeneous architecture that:

- ✓ Enables task offloading on the client side,
- ✓ Handles offloaded tasks in the cloud side with remote CPUs and GPUs,
- ✓ Monitors the activity of physical resources
- ✓ Centralizes the information of resources,
- ✓ and supports the QoS aspects of each user.



Thank you!

www.rapid-project.eu



@RAPID_H2020



Iakovos Mavroidis
jacob@ics.forth.gr

