Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 97 (2016) 118 – 121

Procedia Computer Science

CLOUD FORWARD: From Distributed to Complete Computing, CF2016, 18-20 October 2016, Madrid, Spain

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

Lara López[a,*], Francisco Javier Nieto[a], Terpsichori-Helen Velivassaki[b], Sokol Kosta[c], Cheol-Ho Hong[d], Raffaele Montella[e], Iakovos Mavroidis[f], Carles Fernández[g]

[a]*Atos Spain SA, Albarracin 25, Madrid, 28037, Spain*
[b] *SingularLogic, Al. Panagouli & Siniosoglou, Athens,Greece*
[c]*Aalborg University Copenhagen, Denmark & Sapienza University of Rome, Italy*
[d]*Queen's University Belfast, University Rd, Belfast, United Kingdom*
[e]*University of Napoli Parthenope, CDN Isola C4, Napoli, 80143, Italy*
[f]*Foundation for Research and Technology – Hellas, Heraklion, Greece*
[g]*Herta Security, Pau Claris 165 4B, Barcelona, 08037, Spain*

Abstract

This position paper presents a novel heterogeneous CPU-GPU multi-level cloud acceleration focusing on applications running on embedded systems found on low-power devices. A runtime system performs energy and performance estimations in order to automatically select local CPU-based and GPU-based tasks that should be seamlessly executed on more powerful remote devices or cloud infrastructures. Moreover, it proposes, for the first time, a 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.

Peer-review under responsibility of organizing committee of the international conference on cloud forward: From Distributed to Complete Computing

Keywords: HPC; accelerators; devices; offloading; QoS; Cloud computing

* Corresponding author. Tel.: +34 94 242 1007.
E-mail address: lara.lopez@atos.net

Peer-review under responsibility of organizing committee of the international conference on cloud forward: From Distributed to Complete Computing
doi:10.1016/j.procs.2016.08.287

1. Introduction

Many low-power devices, such as smartphones or tables, as well as several other embedded systems cannot always cope with the increased demand for processing power, memory and storage required by modern applications. As a result, most of these applications are only executed on high-end servers. According to Ericsson Mobility Report[1], there have been 7.3 billion mobile subscriptions in 2015, with an estimation of 5% increase for the next 5 years. Furthermore, the smartphone data traffic is expected to increase by 8.9% in the following 5 years. Users keep demanding more features and services from mobile applications developers. To be able to satisfy users' demands, developers offload the heavy operations of an application to more resourceful machines, such as private or public cloud, giving birth to a new research area known as *Mobile Cloud Computing*[2, 3, 4]. The objective of this paper is to show the approach followed by RAPID[5] project on tackling the above mentioned challenge.

RAPID proposes and designs the full architecture of a heterogeneous offloading framework identifying several components needed for making the system highly automatic and transparent to the developers and final users. Finally, its solution can be applied to many different scenarios, however only three of them will be implemented within the project life: antivirus, hand-tracking and biosurveillance.

Nomenclature

AC	Acceleration Client
API	Application Programming Interface
AS	Acceleration Server
DFE	Dispatch and Fetch Engine
DS	Directory Server
DSE	Design Space Explorer
RM	Registration Manager
SLA	Service Level Agreement
SLAM	Service Level Agreement Manager
VM	Virtual Machine
VMM	Virtual Machine Manager

2. Architecture

The RAPID system architecture (see Fig. 1) consists of five main components: Acceleration Client (AC), Acceleration Server (AS), Directory Server (DS), Service Level Agreement Manager (SLAM), and Virtual Machine Manager (VMM).

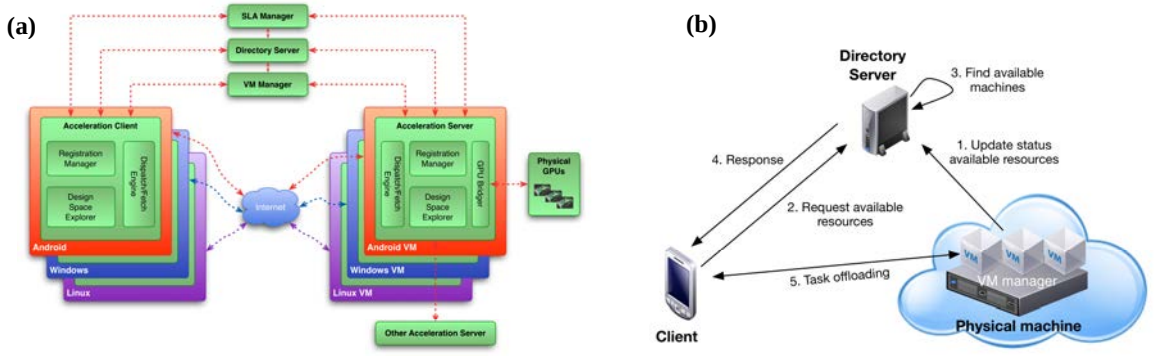

Fig. 1 (a) High-level RAPID logical architecture (b) Interaction between the entities in the RAPID system