LOw Power Heterogeneous Architecture
for NExt Generation of SmaRt Infrastructure and Platforms
in Industrial and Societal Applications

Dissemination – Final Report

Co-funded by the Horizon 2020
Framework Programme of the European Union

DELIVERABLE NUMBER	D8.14
DELIVERABLE TITLE	Dissemination – Final Report
RESPONSIBLE AUTHOR	Idan Yaniv (Technion)

GRANT AGREEMENT N.	688386
PROJECT REF. NO	H2020- 688386
PROJECT ACRONYM	OPERA
PROJECT FULL NAME	LOw Power Heterogeneous Architecture for Next Generation of SmaRt Infrastructure and Platform in Industrial and Societal Applications
STARTING DATE (DUR.)	01/12/2015
ENDING DATE	30/11/2018
PROJECT WEBSITE	www.operaproject.eu

WORKPACKAGE N. | TITLE	WP8 | Exploitation, Impact, dissemination
WORKPACKAGE LEADER	TESEO
DELIVERABLE N. | TITLE	D8.14 | Dissemination – Final Report
RESPONSIBLE AUTHOR	Idan Yaniv (Technion)
DATE OF DELIVERY (CONTRACTUAL)	30/11/2018 (M36)
DATE OF DELIVERY (SUBMITTED)	15/02/2019
VERSION | STATUS	V2.0 | final
NATURE	R (Report)
DISSEMINATION LEVEL	PU (Public)
AUTHORS (PARTNER)	Idan Yaniv (Technion); Alberto Scionti (ISMB); Florin Apopei (TESEO); Joel Nider (IBM)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.1	First draft	14/10/2018	Idan Yaniv (Technion)
0.2	Add the papers published in the last period of the project	14/11/2018	Idan Yaniv (Technion) Florin Apopei (TESEO) Joel Nider (IBM)
0.3	Describe HiPEAC involvement	24/11/2018	Alberto Scionti (ISMB)
0.4	Migrate to the new template	24/11/2018	Idan Yaniv (Technion)
0.5	First internal review	26/11/2018	Joel Nider (IBM)
0.6	Describe some more activities and papers	26/11/2018	Alberto Scionti (ISMB) Richard Chamberlain (Nallatech)
0.7	Second internal review	27/11/2018	Damien Barbot (Neavia)
1.0	Camera-ready version	27/11/2018	Idan Yaniv (Technion)
1.1	Address the reviewers comments	31/01/2019	Idan Yaniv (Technion)
1.2	Add Open Access links to all submissions	31/01/2019	Idan Yaniv (Technion)
2.0	Final review	15/02/2019	Giulio URLINI (ST), Olivier TERZO (ISMB)

PARTICIPANTS		CONTACT
STMICROELECTRONICS SRL		Giulio Urlini Email: Giulio.urlini@st.com
IBM ISRAEL SCIENCE AND TECHNOLOGY LTD		Joel Nider Email: joeln@il.ibm.com
HEWLETT PACKARD CENTRE DE COMPETENCES (FRANCE)		Gallig Renaud Email: gallig.renaud@hpe.com
NALLATECH LTD		Craig Petrie Email: c.petrie@molex.com
ISTITUTO SUPERIORE MARIO BOELLA		Olivier Terzo Email: terzo@ismb.it
TECHNION ISRAEL INSTITUTE OF TECHNOLOGY		Dan Tsafrir Email: dan@cs.technion.ac.il
CSI PIEMONTE		Vittorio Vallero Email: vittorio.vallero@csi.it
NEAVIA TECHNOLOGIES		Stéphane Gervais Email: s.gervais@lacroix.fr
CERIOS GREEN BV		Frank Verhagen Email: frank.verhagen@certios.nl
TESEO SPA		Stefano Serra Email: stefano.serra@eiffage.com
DEPARTEMENT DE L'ISERE		Olivier Latouille Email: olivier.latouille@isere.fr

ACRONYMS LIST

Acronym	Description
EU	EU European Union
WP	Work package
HiPEAC	High Performance and Embedded Architecture and Compilation
HA	Heterogeneity Alliance

LIST OF TABLES

EXECUTIVE SUMMARY

This document---the final dissemination report---summarizes the dissemination efforts that the OPERA consortium carried out in its three years of work. To provide a complete overview of the dissemination results, we will have to introduce again the activities that we already reported. More specifically, we will repeat some parts of the first and second dissemination reports (deliverables 8.8 [2] and 8.12 [3], which were submitted in M16 and M26 of the OPERA project, respectively).

Now that we have arrived in the final stage of the OPERA project, our dissemination activities can be evaluated against the targets that were set in the dissemination plan (initially defined in D8.4 [1] and later enhanced in D8.12 [3]). Table 1 shows that we fully met the goals and even beyond.

Table 1: The OPERA project fully met its dissemination targets.

	Plan	Actual	**Success ratio**
Target #1 (published papers)	9	11	**122%**
Target #2 (researchers reached)	4,000	4,000	**100%**
Target #3 (reports per task)	1	1	**100%**
Target #4 (events/workshops)	10	11	**110%**

The OPERA consortium also enhanced the dissemination activities through the active participation in the events organised by HiPEAC. These events also involved attending sessions and meeting with the Heterogeneity Alliance which have brought us to the preparation of a book (where the OPERA consortium have contributed a chapter). HiPEAC is a well-established community comprising academic, centers of excellence, and industrial members in the domain of high-performance, low-power computer architecture and design (covering both the software and hardware aspects). Attending the HiPEAC events allowed OPERA to create important connections with other Horizon 2020 projects, as well as to share project results.

Position of the deliverable in the whole project context

Dissemination is essential for the success of the project and for the sustainability of outputs in the long term. This report (D8.14) is not only a periodical report of the OPERA dissemination activities but also a summary of all dissemination activities throughout the project by comparing them to the clear, measurable targets that were defined in D8.4 [1] and D8.12 [3]. We believe that this deliverable provides a thorough account of all OPERA activities that were carried out by all partners to distribute information to various audiences within the academic and industrial communities.

This final report demonstrates that the OPERA project has achieved its goals and objectives and, in some cases, even beyond the targets. For example, the OPERA consortium published 10 scientific papers, organized 11 events and workshops. We therefore believe that the OPERA results and findings reached thousands of scientists and engineers.

TABLE OF CONTENTS

1 DISSEMINATION PROGRESS

This chapter describes how the OPERA project disseminated its results and findings in the scientific and industrial communities. We examine the success of the dissemination process against the quantitative targets defined in the OPERA dissemination plan. Note that the original OPERA dissemination plan, which was submitted under Deliverable 8.4 ([1] Dissemination Plan), was significantly enhanced in Deliverable 8.12 ([3] Dissemination Report #2). Henceforth, we will only refer to the enhanced OPERA dissemination plan. The main improvement was to increase the OPERA target for scientific publications from six to nine, according to the comments of the reviewers in the second review (held in Brussels on 20 of July 2017).

Table 2 presents the original dissemination plan and the improved plan. We measure the progress of our dissemination with clear and quantitative targets to allow monitoring and reporting the outcomes:

Target #1: publish 9 papers along the OPERA project in total.

Target #2: reach 4,000 people from the academic community in total. In other words, we want that at least 4,000 researchers will attend the conferences where OPERA works were presented or will read the journals where OPERA works were published.

Target #3: write at least one report for every task.

Target #4: organize 10 workshops/events along the OPERA project in total.

The following sections describe the status of each of the targets, and Table 2 summarizes our progress according to the measurable targets that were set in the dissemination plan. We can see that the OPERA project has met its goals and sometimes even better than expected, e.g., published 10 (and not only 9 as planned) papers in scientific conferences and journals.

Table 2: Progress of the dissemination targets

	Improved Plan D8.12 [3]	Last report (M26)	**Current (M36)**	**Success ratio**
Target #1 (published papers)	9	7	**11**	**122%**
Target #2 (researchers reached)	4,000	3,000	**4,000**	**100%**
Target #3 (reports per task)	1	1	**1**	**100%**
Target #4 (events/workshops)	10	9	**11**	**110%**

1.1 STATUS OF TARGET #1: PUBLICATIONS

The academic and research-oriented partners of OPERA (Technion, ISMB, and IBM) have published four new papers in the last period of the project (M27—M36), as described below. [One of these papers has actually been accepted but not yet published in the next IEEE DATE'19 conference.] These papers and the posters we used to present these works were also published in the OPERA website, the OPERA Facebook page, and the OPERA Twitter page. For the sake of completeness, we provide the entire list of published papers below in a chronological order and describe them in detail, thus repeating some content from previous dissemination reports (D8.4 [2] and D8.12 [3]).

Following the guidelines of EU-funded projects, we also provide Open Access links to all OPERA publications. We emphasize that Open Access is beneficial because it increases the impact and visibility of our research and allows us to share our results such that money is no object.

1.1.1 Paper #1: Hash, Don't Cache (The Page Table)

Authors: Idan Yaniv, Dan Tsafrir.

Title: "Hash, Don't Cache (the Page Table)".

Publication type: conference paper (peer reviewed).

Venue: ACM SIGMETRICS International Conference on Measurement and Modeling (SIGMETRICS).

Date: June 2016.

Location: Antibes Juan-les-Pins, France.

DOI: https://doi.org/10.1145/2896377.2901456

Open Access Link: http://www.operaproject.eu/wp-content/uploads/2016/11/hvsr.pdf

Related to: WP4, Task 4.2.

Abstract:

Radix page tables as implemented in the x86-64 architecture incur a penalty of four memory references for address translation upon each TLB miss. These 4 references become 24 in virtualized setups, accounting for 5%–90% of the runtime and thus motivating chip vendors to incorporate page walk caches (PWCs). Counterintuitively, an ISCA 2010 paper found that radix page tables with PWCs are superior to hashed page tables, yielding up to 5x fewer DRAM accesses per page walk. We challenge this finding and show that it is the result of comparing against a suboptimal hashed implementation—that of the Itanium architecture. We show that, when carefully optimized, hashed page tables in fact outperform existing PWC-aided x86-64 hardware, shortening benchmark runtimes by 1%–27% and 6%–32% in bare-metal and virtualized setups, without resorting to PWCs. We further show that hashed page tables are inherently more scalable than radix designs and are better suited to accommodate the ever increasing memory size; their downside is that they make it more challenging to support such features as superpages.

1.1.2 Paper #2: Workload Management for Power Efficiency in Heterogeneous Data Centers

Authors: Pietro Ruiu, Alberto Scionti, Joel Nider, Mike Rapoport.

Title: "Workload Management for Power Efficiency in Heterogeneous Data Centers".

Publication type: conference paper (peer reviewed).

Venue: International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS).

Date: July 2016.

Location: Fukuoka, Japan.

DOI: https://doi.org/10.1109/CISIS.2016.107

Open Access Link: http://www.operaproject.eu/wp-content/uploads/2016/11/OPERA_CISIS2016.pdf

Related to: WP5, all tasks.

Abstract:

The cloud computing paradigm has recently emerged as a convenient solution for running different workloads on highly parallel and scalable infrastructures. One major appeal of cloud computing is its

capability of abstracting hardware resources and making them easy to use. Conversely, one of the major challenges for cloud providers is the energy efficiency improvement of their infrastructures. Aimed at overcoming this challenge, heterogeneous architectures have started to become part of the standard equipment used in data centers. Despite this effort, heterogeneous systems remain difficult to program and manage, while their effectiveness has been proven only in the HPC domain. Cloud workloads are different in nature and a way to exploit heterogeneity effectively is still lacking. This paper takes a first step towards an effective use of heterogeneous architectures in cloud infrastructures. It presents an in-depth analysis of cloud workloads, highlighting where energy efficiency can be obtained. The microservices paradigm is then presented as a way of intelligently partitioning applications in such a way that different components can take advantage of the heterogeneous hardware, thus providing energy efficiency. Finally, the integration of microservices and heterogeneous architectures, as well as the challenge of managing legacy applications, is presented in the context of the OPERA project.

1.1.3 Paper #3: OPERA: A Low Power Approach to the Next Generation Cloud Infrastructures

Authors: Alberto Scionti, Pietro Ruiu, Olivier Terzo, Joel Nider, Craig Petrie, Niccolo Baldoni.

Title: “OPERA: a Low Power Approach to the Next Generation Cloud Infrastructures”.

Publication type: conference paper (peer reviewed).

Venue: Euromicro Conference on Digital System Design (DSD).

Date: August 2016.

Location: Limassol, Cyprus.

DOI: https://doi.org/10.1109/DSD.2016.63

Open Access Link: http://www.operaproject.eu/wp-content/uploads/2019/01/07723570.pdf

Related to: WP2, WP3, WP6, WP7.

Abstract:

The continuous evolution of information and communication technology has led to a change in the adopted computing paradigms over time. Cloud computing is an emerging paradigm in which users, depending on their specific requirements, access to a shared pool of computing resources dynamically allocated. Cloud computing represents, with respect to Grid computing, the evolutionary step towards the implementation of a ubiquitous computing service. Such paradigm leverages on the infrastructural capabilities (compute, storage, and network) of modern data centers to provide an adequate level of computational power able to satisfy users’ requests. However, trying to continuously increase such capabilities comes at the cost of an increased energy consumption. Energy efficiency is, therefore, one of the major challenges that cloud providers must address. The OPERA project aims at bringing innovative solutions to increase the energy efficiency of cloud infrastructures, by leveraging on modular, high-density, heterogeneous and low power computing systems, which are able to cover the whole computing continuum. To this end, the project will design a highdensity server solution in which low power processors and FPGA devices will be used to accelerate cloud workloads. High-speed optical interconnections will be used to connect the proposed server with high-performance nodes, such as OpenPOWER-based machines. Cyber-Physical Systems (CPS) represents a natural extension of cloud infrastructures since they can collect and process data locally, more specifically where they were generated. OPERA aims at researching energy efficiency of such cloud endnodes by designing an ultra-low power computing system with reconfigurable radio frequency capabilities. The effectiveness of the whole platform will be demonstrated with key scenarios, specifically a road traffic monitoring application, the deployment of a virtual desktop infrastructure, and the deployment of a small data

center on a truck.

1.1.4 Paper #4: Remote page faults with a CAPI based FPGA

Authors: Nider, J., Binyamini, Y., Rapoport, M.

Title: “Remote page faults with a CAPI based FPGA”.

Publication type: poster paper (peer reviewed).

Venue: ACM International Systems and Storage Conference (SYSTOR).

Date: May 22 – 24, 2017.

Location: Haifa, Israel.

DOI: https://doi.org/10.1145/3078468.3078489

Open Access Link: http://www.operaproject.eu/wp-content/uploads/2017/07/a20-nider.pdf

Related to: WP5, Task 5.2.

Abstract:

Post-copy VM or container migration requires that the bulk of the memory is transferred after resuming on the destination node. Transferring memory between nodes over a commodity TCP/IP network incurs too much latency, which slows down execution of the application on the destination node.

1.1.5 Paper #5: User space memory management for post-copy migration

Authors: Rapoport, M., Nider, J.

Title: “User space memory management for post-copy migration”.

Publication type: poster paper (peer reviewed).

Venue: ACM International Systems and Storage Conference (SYSTOR).

Date: May 22 – 24, 2017.

Location: Haifa, Israel.

DOI: https://doi.org/10.1145/3078468.3078490

Open Access Link: http://www.operaproject.eu/wp-content/uploads/2017/07/a21-rapoport.pdf

Related to: WP5, Task 5.2.

Abstract:

Post-copy migration allows reduction of application down-time and reduces overall network bandwidth used for application migration. Migration can be used to help optimize several aspects of operations such as power efficiency. The userfault technology recently introduced to the Linux kernel allows post-copy migration of virtual machines. However, this technology is missing essential features required for post-copy migration of Linux containers.

1.1.6 Paper #6: Systems Software for Fast Inter-Machine Page Faults

Authors: Nider, J., Rapoport, M., Binjamini, Y.

Title: “Systems Software for Fast Inter-Machine Page Faults”.

Publication type: conference paper (peer reviewed).

Venue: Future Technologies Conference (FTC).

Date: November 29-30, 2017.

Location: Vancouver, Canada.

DOI: N/A.

Open Access Link: http://www.operaproject.eu/wp-content/uploads/2019/01/102_Paper_446-Systems_Software_for_Fast_Inter-Machine_Page_Faults.pdf

Related to: WP5, Task 5.2.

Abstract:

Cloud computing abstracts the underlying hardware details from the user. As long as the customer Service Level Agreements (SLA) are satisfied, cloud providers and operators are free to make infrastructural decisions to optimize business objectives, such as operational efficiency of cloud data centers. By adopting a holistic view of the data center and treating it as a single system, a cloud provider can migrate application components and virtual machines within the system according to policies such as load balancing and power consumption. We contribute to this vision by removing architectural barriers for workload migration and reducing the downtime of migrating processes. We combine the post-copy approach to workload migration with a novel specialized low latency interconnect for handling resulting remote page faults. In this work, we introduce a cross-architecture workload migration system, specify the requirements towards the specialized interconnect, discuss design trade-offs and issues, and present our proposed SW-HW co-design.

1.1.7 Paper #7: Open-source implementation of an Ad-hoc IEEE 802.11 a/g/p software-defined radio on low-power and low-cost general-purpose processors

Authors: Ciccia, S., G. Giordanengo, and G. Vecchi.

Title: “Open-source implementation of an Ad-hoc IEEE 802.11 a/g/p software-defined radio on low-power and low-cost general purpose processors”.

Publication type: journal article (peer reviewed).

Venue: Radioengineering journal, volume 26, number 4, pages 1083—1095.

Date: December 2017.

DOI: http://dx.doi.org/10.13164/re.2017.1083

Open Access Link: http://www.operaproject.eu/wp-content/uploads/2017/12/17_04_1083_1095.pdf

Related to: WP3, Task 3.4.

Abstract:

This work proposes a low-cost and low-power software-defined radio open-source platform with IEEE 802.11 a/g/p wireless communication capability. A state-of-the-art version of the IEEE 802.11 a/g/p software for GNU Radio (a free and open-source software development framework) is available online, but we show here that its computational complexity prevents operations in low-power general purpose processors, even at throughputs below the standard. We therefore propose an evolution of this software that achieves a faster and lighter IEEE 802.11 a/g/p transmitter and receiver, suitable for low-

power general purpose processors, for which GNU Radio provides very limited support; we discuss and describe the software radio processing structuring that is necessary to achieve the goal, providing a review of signal processing techniques. In particular, we emphasize the advanced reduced-instruction set (RISC) machine (ARM) study case, for which we also optimize some of the processing libraries. The presented software will remain open-source.

1.1.8 Paper #8: Towards Energy Efficient Orchestration of Cloud Computing Infrastructure

Authors: A. Scionti, K. Goga, F. Lubrano, O. Terzo.

Title: "Towards Energy Efficient Orchestration of Cloud Computing Infrastructure"

Publication type: conference paper (peer reviewed).

Venue: Complex, Intelligent, and Software Intensive Systems (CISIS).

Date: 4—6 July 2018.

Location: Matsue, Japan.

DOI: https://www.doi.org/10.1007/978-3-319-93659-8_15

Open Access Link: https://link.springer.com/chapter/10.1007%2F978-3-319-93659-8_15

Related to: WP2, WP5.

Abstract:

The emerging of new Cloud services and applications demanding for ever more performance (i.e., on one hand, the rapid growth of applications using deep learning –DL, on the other hand, HPC-oriented work-flows executed in Cloud) is continuously putting pressure on Cloud providers to increase capabilities of their large data centers, by embracing more advanced and heterogeneous devices. Hardware heterogeneity also helps Cloud providers to improve energy efficiency of their infrastructures by using architectures dedicated to specific workloads. However, heterogeneity represents a challenge from the infrastructure management perspective. In this highly dynamic context, workload orchestration requires advanced algorithms to not defeat the efficiency provided by the hardware layer. Despite past works partially addressed the problem, a comprehensive solution is still missing.

This paper presents the solution studied within the European H2020 project OPERA. Our approach is intended for managing the workload in large infrastructures running heterogeneous systems, by using a two-steps approach. Whenever new jobs are submitted, an energy-aware allocation policy is used to select the most efficient nodes where to execute the incoming jobs. In a second step, the whole workload is consolidated by means of the optimization of a cost model. This paper focuses on an allocation algorithm aimed at reducing the overall energy consumption; it also presents the results of simulations on a State-of-the-Art framework. When compared with well-known and broadly adopted allocation strategies, the proposed approach results in a tangible energy-saving (up to 30% compared to First Fit allocation policy, and up to 45.2% compared to the Best Fit), thus demonstrating energy efficiency superiority.

1.1.9 Paper #9: The Potential Influence of Workload Management Across Heterogeneous Server Systems on Datacenter Energy Use and Power Draw

Authors: A. Scionti, D. Harryvan, R. Chamberlain, O. Terzo, G. Urlini.

Title: "The Potential Influence of Workload Management Across Heterogeneous Server Systems on Datacenter Energy Use and Power Draw"

Publication type: conference paper (peer reviewed).

Venue: International Telecommunications Energy Conference (INTELEC).

Date: 7—11 October 2018.

Location: Turin, Italy

DOI: https://doi.org/10.1109/INTLEC.2018.8612307

Open Access Link: http://www.operaproject.eu/wp-content/uploads/2019/01/08612307.pdf

Related to: WP5, all tasks.

Abstract:

Abstract—Energy efficiency is a key part of the European energy policies and 2020 climate targets. Project OPERA is working to create an energy aware workload manager for heterogeneous systems that will allow microservices to migrate between systems with differing instruction set architectures. The Energy savings potential of such technologies is enormous and is estimated at 47 TWh per year in Europe, 95% of the energy consumed by servers in Europe. The impact of such technologies on datacenter operations is profound. Significant and fast variations in power draw over time are expected, a fact that operators need to consider when retrofitting or designing new facilities.

1.1.10 Paper #10: Empty Promise: Zero-copy Receive for vhost

Authors: Kalman Meth, Mike Rapoport, Joel Nider

Title: "Empty Promise: Zero-copy Receive for vhost"

Publication type: conference paper (peer reviewed).

Venue: Future Technologies Conference (FTC).

Date: 13-14 November 2018.

Location: Vancouver, BC, Canada.

DOI: https://doi.org/10.1007/978-3-030-02683-7_45

Open Access Link: https://link.springer.com/chapter/10.1007%2F978-3-030-02683-7_45

Related to: WP5, Task 5.2.

Abstract:

When receiving network traffic on guest VMs (virtual machines) running on the KVM hypervisor, data is copied from hypervisor buffers to guest buffers. This copy incurs CPU and latency overhead. It seems desirable to avoid this copy, if possible, and to use guest buffers directly to receive network data. A number of challenges must be overcome in order to accomplish this zero-copy receive, and some additional overhead is incurred in setting it up. We present the technical details on a zero-copy receive design and implementation and discuss the resulting performance degradation. In our implementation, the savings in CPU from avoiding the copy is overshadowed by the extra handling required for the smaller and often underutilized buffers.

1.1.11 Paper #11: Chip-to-Cloud: an Autonomous and Energy Efficient Platform for Smart Vision Applications

Authors: Alberto Scionti, Simone Ciccia, Olivier Terzo, Giorgio Giordanengo

Title: “Chip-to-Cloud: an Autonomous and Energy Efficient Platform for Smart Vision Applications”

Publication type: conference paper (peer reviewed).

Venue: Design, Automation & Test in Europe Conference & Exhibition (DATE).

Date: TBD (25-29 March 2019).

Location: Firenze, Italy.

DOI: N/A.

Open Access Link: will be provided after the paper is published.

Related to: WP3, WP7.

Abstract:

Modern Cloud architectures encompass computing and communication elements that span from traditional data center computing nodes to edge-computing and IoT devices. Modern data centers offer almost infinite resources to satisfy any application demands; while, edge-computing and IoT extend Cloud platforms to sense and act on the real world (Cyber-Physical Systems –CPS). This paper presents the Cloud architecture devised within the OPERA project, which provides new levels of energy efficiency as a full chip-to-Cloud solution. Focusing on a smart vision application (i.e., road traffic monitoring), the paper presents novel architectural solutions optimised to achieve high energy efficiency at any level: *i*) the computing elements supporting the acceleration of State-of-the-Art CNNs; and *ii*) an innovative wireless communication subsystem. Unlike conventional designs, our wireless communication subsystem exploits the advantages of software defined radio (SDN) firmware to control a reconfigurable antenna. To further extend CPS application range, an energy harvesting module is used to supply power. Besides the CPS, high-density accelerated servers offer capabilities of running complex algorithms within a small power envelop. The effectiveness of the whole architecture has been tested in a real context (i.e., 2 installation sites). In-field measurements demonstrate our claim: high-performance coupled with high energy efficiency over the whole system.

1.1.12 Paper #12: The Green Computing Continuum: The OPERA Perspective

Authors: A. Scionti, O. Terzo, P. Ruiu, G. Giordanengo, S. Ciccia, G. Urlini, J. Nider, M. Rapoport, C. Petrie, R. Chamberlain, G. Renaud, D. Tsafrir, I. Yaniv and D. Harryvan.

Title: “The Green Computing Continuum: The OPERA Perspective”

Publication type: book chapter.

Venue: the book “Hardware Accelerators in Data Centers”, published by Springer, Cham.

Date: 22 August 2018.

DOI: https://doi.org/10.1007/978-3-319-92792-3_4

Open Access Link: https://link.springer.com/chapter/10.1007%2F978-3-319-92792-3_4

Related to: all WPs.

Abstract:

Cloud computing is an emerging paradigm in which users' access to a shared pool of computing resources is dynamically allocated (i.e. ubiquitous computing service), depending on their specific needs. Such paradigm exploits the infrastructural capabilities of modern data centers to provide computational power and storage space required to satisfy modern application demands. The seamless integration of Cyber-Physical Systems (CPS) and Cloud infrastructures allows the effective processing of the huge amount of data collected by smart embedded systems, towards the creation of new services for the end users. However, trying to continuously increase data center capabilities comes at the cost of an increased energy consumption. The OPERA project aims at bringing innovative solutions to increase the energy efficiency of Cloud infrastructures, by leveraging on modular, high-density, heterogeneous and low-power computing systems, spanning data center servers and remote CPS. The effectiveness of the proposed solutions is demonstrated with key scenarios: a road traffic monitoring application, the deployment of a virtual desktop infrastructure, and the deployment of a compact data center on a truck.

1.2 STATUS OF TARGET #2: COMMUNITY REACH

According to the communication reports filled by all OPERA partners, the estimated number of researchers who were exposed to the OPERA findings and products along the lifetime of the project (M1—M36) is 4,000. Some of these are scientists and engineers who attended the conferences where OPERA works were presented or read the journals where OPERA works were published. Other researchers have heard about the OPERA innovative plans and products through presentations and posters delivered by the OPERA partners. We list some of the dissemination activities that happened in the last period of the project (M27—M36) below:

- IBM: Mike Rapoport presented the post-copy migration work at the Free and Open Source Software Developers' European Meeting (FOSDEM) 2018 at Brussels, Belgium. A video of Mike's FOSDEM talk can be watched in: https://www.youtube.com/watch?v=8ADMmrhGd-M
- Technion: Mohammad Agbarya and Idan Yaniv presented their research under WP5 at the ACM International Systems and Storage Conference (SYSTOR) 2018. Their poster can be found in: https://www.systor.org/2018/pdf/systor18-19.pdf
- IBM: Mike Rapoport and Joel Nider presented both the post-copy container migration work as well as the RDMA remote page fault work at the Linux Plumber's Conference (LPC) 2018 in Vancouver, Canada. The slides can be found on the Linux Plumber's site: https://linuxplumbersconf.org/event/2/contributions/205/ and https://linuxplumbersconf.org/event/2/contributions/135/
- Nallatech: Richard Chamberlain attended several conferences: ISC2017, SC2017, ISC2018 (Germany) and SC2018 (US) computing conferences and presented OPERA dissemination material on the Nallatech exhibition booth. This included white papers regarding the CNN offload process, that were updated in line with the progress of the project. Nallatech also presented OPERA progress at the HPE TES events in France and computing insight conferences in the UK.

1.3 STATUS OF TARGET #3: TASK REPORTING

In the original OPERA grant agreement, all OPERA tasks committed to deliver one document that summarizes the work under this task. According to the feedback from the reviewers in the first review (M9), most of the tasks further committed to one or more intermediate reports and deliverables, thus enhancing and promising the achievement of Target #3.

1.4 STATUS OF TARGET #4: ORGANIZING WORKSHOPS

In the context of dissemination activities, the OPERA project was very active within all the initiatives organised by the HiPEAC European network of excellence. The HiPEAC was founded as a network of excellence in the field of "High Performance and Embedded Architecture and Compilation", and during several years provided spaces for both academic and industrial communities to share their ideas.

Since 2016, the OPERA project is an active member of the network, not only attending all the events organised, but also organising by itself workshops held and co-located with the main events. As such, for the last reporting period of the project (M27—M36), OPERA attended both the main conference (January 2018 -- Manchester, UK) and the Computing System Weeks (CSWs -- May 2018, Gothenburg, S, and October 2018, Heraklion, GR). These events represented the opportunity to organize a dedicated thematic session and a workshop. Both the thematic session (i.e., a workshop dedicated to a specific topic) and the workshop (named HeLP-DC -- Heterogeneous and Low Power technologies for Data Centers) gave the OPERA project the opportunity to disseminate project results, reaching a large number of people in the academic and industrial community. Also, other EU projects have been invited to share their results, thus resulting in a positive and active share of strength points and solutions designed by OPERA project.

This sharing also contributed to enable more strict collaborations between funded EU projects sitting in the same domain (e.g., M2DC, VINEYARD, etc.). It is worth to mention that HiPEAC events provided the space for sharing ideas, solutions and project results with the Heterogeneity Alliance (HA) members (also, OPERA is member of such initiative), which brought to setting up the writing of a technical book, where OPERA contributed. Similarly to HiPEAC, HA is born under the umbrella of a EU project (TANGO) and quickly grew by including other EU projects and institutions as full active members. To this end, results of the OPERA project have been shared with HA.

2 CONCLUSIONS

This deliverable surveys the OPERA dissemination plan and its outcomes in terms of quantitative goals. The dissemination plan was initially outlined in D8.4 [1] and was later extended in D8.12 [3] according to the recommendations of the second review report, notably by increasing the number of scientific publications. We carefully report the dissemination efforts that were carried out in the last 10 months (M27—M36) of the project. Following the guidelines of EU-funded projects, we also provide Open Access links to all OPERA publications, thereby increasing the impact and visibility of our research.

Last, we described the HiPEAC-related events that we organised by the OPERA consortium to ensure that the project results will be distributed to various audiences within the European academic and research-oriented community. HiPEAC is a well-established community comprising academic, centers of excellence, and industrial members in the domain of high-performance, low-power computer architecture and design. Besides sharing the OPERA project results, attending the HiPEAC events also allowed OPERA to create important connections with other Horizon 2020 projects.

ATTACHMENTS

REFERENCES

[1] D8.4, Dissemination Plan
[2] D8.8, Dissemination Report 1
[3] D8.12, Dissemination Report 2