THE 13TH CONFERENCE OF PHD STUDENTS IN

COMPUTER SCIENCE

Volume of short papers

CS^2

Organized by the Institute of Informatics of the University of Szeged

June 29 – July 1, 2022
Szeged, Hungary

Scientific Committee:

János Csirik (Co-Chair, SZTE)
Lajos Rónyai (Co-Chair, SZTAKI, BME)
András Benczúr (ELTE)
András Benczúr ifj. (SZTAKI)
Tibor Csendes (SZTE)
László Cser (Corvinus University of Budapest)
Erzsébet Csuhaj-Varjú (ELTE)
József Dombi (SZTE)
Zoltán Fülöp (SZTE)
Aurél Galántai (Óbuda University)
Zoltán Gingl (SZTE)
Tibor Gyimóthy (SZTE)
Katalin Hangos (MTA)
Zoltán Horváth (ELTE)
Márk Jelasity (SZTE)
Tibor Jordán (ELTE)
Zoltán Kása (Sapientia University)
Zoltán Kató (SZTE)
László T. Kóczy (BME)
Andrea Kő (Corvinus University of Budapest)
Miklós Kuczmann (SZE)
János Levendovszky (BME)
Gyöngyvér Márton (Sapientia University)
Branko Milosavljević (University of Novi Sad)
Valerie Novitzka (Technical University of Kosice)
Laszló G. Nyúl (SZTE)
Marius Otesteanu (Politehnica University Timisoara)
Attila Pethő (University of Debrecen)
Sándor Radeleczki (University of Miskolc)
András Recski (BME)
Sándor Szabó (PTE)
Gábor Szederkényi (PPKE)
János Sztrik (University of Debrecen)
János Tapolcai (BME)
József Tar (Óbuda University)
Gyula Vastag (Corvinus University of Budapest)
János Végh (University of Miskolc)

Organizing Committee:
Judit Jász, Balázs Bánhelyi, Tamás Gergely, Melinda Katona, Zoltán Kincses
Address of the Organizing Committee
c/o. Judit Jász
University of Szeged, Institute of Informatics
H-6701 Szeged, P.O. Box 652, Hungary
Phone: +36 62 546 728, Fax: +36 62 546 397
E-mail: cscs@inf.u-szeged.hu
URL: http://www.inf.u-szeged.hu/~cscs/

Sponsors

Regional Committees of the
Hungarian Academy of Sciences

University of Szeged,
Institute of Informatics

Preface

This conference is the 13th in a series. The organizers aimed to bring together PhD students working on any field of computer science and its applications to help them publishing one of their first papers, and provide an opportunity to hold a scientific talk. As far as we know, this is one of the few such conferences. The aims of the scientific meeting were determined on the council meeting of the Hungarian PhD Schools in Informatics: it should

- provide a forum for PhD students in computer science to discuss their ideas and research results;
- give a possibility to have constructive criticism before they present the results at professional conferences;
- promote the publication of their results in the form of fully refereed journal articles; and finally,
- promote hopefully fruitful research collaboration among the participants.

The papers emerging from the presented talks will be invited to be considered for full paper publication the Acta Cybernetica journal.

Szeged, June 2022

Judit Jász
Balázs Bánhelyi
Tamás Gergely
Melinda Katona
Zoltán Kincses

Contents

Towards Modelling IoT Workflows

András Márkus

Abstract: The cooperation of distributed computing and Internet of Things (IoT) paradigms created numerous research challenges. Modern applications do not only compute certain tasks, but they support various events that make human life more colourful. Considering workflows in the IoT domain, opposed to general scientific workflows, the focus is on the optimal execution of predefined sequences of various steps, however, computational tasks can also be contained. Such a sequence of steps usually consists of performing a service call, receiving a data packet in the form of a message sent by an IoT device, or executing a computational task on a virtual machine. The development and testing of such IoT workflows and their management systems in real life can be complicated due to high costs and access limitations, therefore simulation solutions should be preferred. In this paper, we discuss the current capabilities of scientific workflow simulation environments, define the needs of IoT workflow modelling support, and make suggestions for a future realisation.

Keywords: Workflow, Internet of Things, Fog Computing, Simulation

Introduction

The advancement of distributed computing paradigms gave birth to various domains such as Grid Computing, Cloud Computing and Fog Computing. Scientific computations leverage this technological involvement, and instead of utilising a monolithic application, data processing can be done in a distributed and scalable fashion exploiting virtualisation and the geographically distributed nature of these environments. In early times, scientific workflows such as Montage, Epigenomics, LIGO, CyberShake or SIPHT from the field of astronomy, physics and bioinformatics were executed on Grid environments [1]. The on-demand resource provisioning, virtualisation and accessibility made available by Cloud Computing opened ways for evolution, however researchers still need to solve optimisation problems of resource utilisation and task scheduling on virtual machines [2].

These types of grid and cloud applications can be considered as workflows described with a directed acyclic graph (DAG). In a DAG, vertices represent a task (i.e. job), whilst edges define the dependencies among the tasks. Until the dependencies of a task are not resolved, it must be considered as suspended. The number of incoming and outgoing edges is not restricted to one, except the start and stop events where in and out directions are missing, respectively. An example of such a scientific workflow is depicted in Figure 1.

Involving the Internet of Things (IoT) domain, IoT workflows must meet novel requirements. Billions of smart devices, gadgets and sensors generate vast amounts of data, therefore IoT is often paired to Fog Computing to enhance data processing and minimise latency by connecting to computing nodes located close to end-users. IoT-Fog-Cloud systems may cover many aspects of our daily life in the frame of smart homes, healthcare and autonomous driving, which typically follow the sense-process-actuate model. IFTTT [3] offers trigger-based online services for IoT, for instance if a new sleep is logged to a smartwatch then a service modifies the room temperature accordingly. Such IFTTT applications or service compositions can be considered as IoT workflows.

Currently there are only a few simulation tools and software solutions aiming execution of workflows in Cloud and Fog environments, however the existing, general workflow description formats are unable to define similar IoT workflows, we defined earlier. To the best of our knowledge, the literature lacks unified IoT workflow description and simulation tools, thus this work represent the first step in this direction.

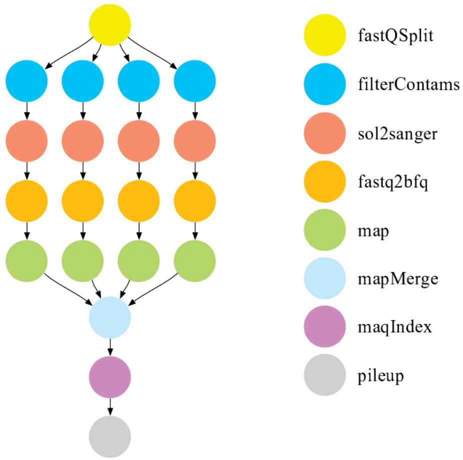

Figure 1: Epigenomics workflow [4]

Research background

Workflow description languages are not executable themselves, therefore typically require execution engines to be run. For scientific workflows (e.g. Montage) the XML-based DAX[1] description is applied. DAX is typically used by Pegasus [1], which was designed to map abstract workflows onto Grids. CWL[2] (Common Workflow Language) can be written in JSON or YAML to define command line tools and corporate them in order to create workflows. OpenWDL[3] (Open Workflow Description Language) is a human readable and writable way to express tasks and workflows, it has an dedicated, unique grammar. GCP (Google Cloud Platform)[4] also has workflow support accepting JSON or YAML-based definitions. It is dedicated to combining and executing GCP services and custom services hosted on GCP.

Simulation is one of the most accepted ways to design and evaluate IoT-Fog-Cloud systems as these tools are capable of modelling such complex systems realistically. The literature is rich with various simulation environments, however only a few open-source tools exist especially dedicated to workflow modelling.

Table 1: Comparison of the related simulators

Simulator	Programming language	Last updated	Dependency	Trace support
WorkflowSim	Java	2015	CloudSim	DAX
FogWorkflowSim		2019	iFogSim, WorkflowSim	
EdgeWorkflowReal		2021	FogWorkflowSim	

WorkflowSim [5] is an extension of the well-known CloudSim simulator written in Java. CloudSim has no workflow support by default, therefore the extended version provides numerous modules for workflow management. The loader module only supports importing DAG files formatted in XML (i.e. DAX), thus the primer goal was to simulate the behaviour of the previously mentioned scientific workflows (e.g. Montage) in distributed cloud environments. The main contribution is the horizontal clustering, which merges tasks into jobs so as to decrease the scheduling overheads. FogWorkflowSim [6] is also a broadening CloudSim-based tool dedicated to model Fog Computing called iFogSim. Some components are also built upon

[1]DAX description (visited on 28.03.2022): https://pegasus.isi.edu/documentation/development/schemas.html
[2]CWL description (visited on 28.03.2022): https://www.commonwl.org/
[3]OpenWDL description (visited on 28.03.2022): https://openwdl.org/
[4]GCP description (visited on 28.03.2022): https://cloud.google.com/workflows

WorkflowSim. FogWorkflowSim supports many greedy and GA-based workflow scheduling algorithms by default, but unfortunately IoT-related workflow simulations are totally omitted. EdgeWorkflowReal [7] inherited numerous function of FogWorkflowSim, furthermore it is capable of creating a real edge computing environment to execute workflow tasks and compare to the simulated results. It also uses a database to store the structure of the real workflow and the data for workflow tasks. Even if this work aims to simulate edge workflow applications, only scientific workflows (i.e. Montage) were considered for the evaluation. The comparison of the discussed simulators can be seen in Table 1.

According to the summarised research background, our observations are as follows: (i) though the discussed simulation tools cover the domains of distributed computing paradigms, IoT specific behaviour is not considered at all. To define a task in a scientific workflow, the length of the task and the size of the incoming/outgoing files (i.e. dependent tasks) are enough to be known. However, typical IoT-related tasks require more freely configurable options, for instance waiting for a message (i.e. sensor data), notifying an actuator entity or expecting some kind of human interactions. It is clearly seen that there is a very strong software dependency (ii) among the different simulation approaches, which could make the modification towards IoT workflows harder, especially if some of the tools were updated years ago. DAX is obviously the leading format applied by simulation communities, due to transparency and shorter learning curve, so it might be a reasonable choice as the used description format (iii) for IoT workflows as well.

Proposed solution

As discussed earlier, IoT workflows differ from the common scientific workflows, because additional factors should be taken into account. To distinguish and highlight their differences and similarities, we introduce an office routine IoT workflow, as shown in Figure 2. As a similarity, the dependencies between tasks still exist, for instance until every sensor sends one package of data and it arrives at the computing node, the data processing cannot be started. As a consequence, computing tasks remain in the system as in case of scientific workflows. As a novelty, time-dependent triggering events also appear in the system. For example, when a user (programmer in this case) enters the office, a service executes an actualisation event by sending instruction to an actuator to start an event automatically (e.g. turning on/off AC, lights or start brewing coffee). Furthermore, time-dependent, recurring events are also part of the system, e.g. measure temperature and humidity in every 30 seconds in the office, until the programmer checks out.

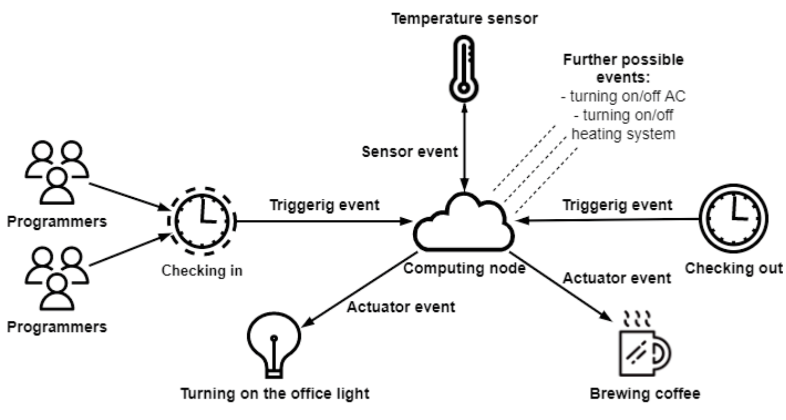

Figure 2: Office routine workflow

Such complex IoT workflows mostly cannot be investigated on a very large scale, when the number of participant entities exceeds thousands in real life, due to the cost implications and the need for large-scale behaviour. Therefore, in the future we plan to develop a simulation extension of DISSECT-CF-Fog, which aims to model independent IoT workflow simulations besides the general scientific workflows. DISSECT-CF-Fog is able to model various IoT device and application behaviours, as well as data centre management with realistic network settings. It also considers energy consumption, IoT device mobility, and pricing schemes of real providers. According to the preliminary plans, three extra modules should be implemented: (i) abstract workflow handler, which transform the human readable format into tasks (ii) task manager is responsible to map the tasks to the proper executors according to the type of a task and finally, (iii) monitoring and reporting step analyses the results.

As a conclusion, in this paper we outlined the current position of workloads, simulation environments and description formats regarding distributed computing environments and we also discussed a research gap and a possible simulation solution in the field of IoT workflows.

Acknowledgements

The research leading to these results is supported by the Hungarian Scientific Research Fund under the grant number OTKA FK 131793, the UNKP-21-3 New National Excellence Program of the Ministry for Innovation and Technology, and by the National Research, Development and Innovation Office within the framework of the Artificial Intelligence National Laboratory Programme.

References

[1] E. Deelman, J. Blythe, Y. Gil, C. Kesselman, G. Mehta, S. Patil, M. Su, K. Vahi amd M. Livny, Miron. Pegasus: Mapping Scientific Workflows onto the Grid. Grid Computing, pp. 11-20, 2004. DOI: 10.1007/978-3-540-28642-4_2

[2] R. Haidri, C.P. Katti and P. Saxena. Cost Effective Deadline Aware Scheduling Strategy for Workflow Applications on Virtual Machines in Cloud Computing. Journal of King Saud University - Computer and Information Sciences, Volume 32, 2017. DOI: 10.1016/j.jksuci.2017.10.009

[3] X. Mi, F. Qian, Y. Zhang and X. Wang. An empirical characterization of IFTTT: ecosystem, usage, and performance. Internet Measurement Conference, pp. 398-404, 2017. DOI: 10.1145/3131365.3131369

[4] S. Bharathi, A. Chervenak, E. Deelman, G. Mehta, M.-H. Su and K. Vahi. Characterization of scientific workflows. Third Workshop on Workflows in Support of Large-Scale Science, pp. 1-10, 2008. DOI: 10.1109/WORKS.2008.4723958

[5] W. Chen, and E. Deelman. WorkflowSim: A toolkit for simulating scientific workflows in distributed environments. International Conference on E-Science, pp. 1-8, 2012. DOI: 10.1109/eScience.2012.6404430

[6] X. Liu, L. Fan, J. Xu, X. Li, L. Gong, J. Grundy and Y. Yang. FogWorkflowSim: an automated simulation toolkit for workflow performance evaluation in fog computing. IEEE/ACM International Conference on Automated Software Engineering, pp. 1114-1117, 2019. DOI: 10.1109/ASE.2019.00115

[7] X. Li, R. Ding, X. Liu, J. Xu, Y. Yang and J. Grundy. EdgeWorkflowReal: An Edge Computing based Workflow Execution Engine for Smart Systems. arXiv preprint, 2021. DOI: 10.48550/arXiv.2102.00234