



HAL
open science

Keynote Talk: Leveraging the Edge-Cloud Continuum to Manage the Impact of Wildfires on Air Quality

Daniel Balouek, Ismael Perez, Sam D Faulstich, Heather A Holmes, Ilkay
Altintas, Manish Parashar

► To cite this version:

Daniel Balouek, Ismael Perez, Sam D Faulstich, Heather A Holmes, Ilkay Altintas, et al.. Keynote Talk: Leveraging the Edge-Cloud Continuum to Manage the Impact of Wildfires on Air Quality. PerCom 2023 - IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events, Mar 2023, Atlanta, United States. pp.27-31, 10.1109/PerComWorkshops56833.2023.10150251 . hal-04387579

HAL Id: hal-04387579

<https://inria.hal.science/hal-04387579v1>

Submitted on 11 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Keynote Talk: Leveraging the Edge-Cloud Continuum to Manage the Impact of Wildfires on Air Quality

Daniel Balouek-Thomert^{*†}, Ismael Perez[‡], Sam D. Faulstich[†], Heather A Holmes[†], Ilkay Altintas[‡], Manish Parashar[†]

^{*}IMT Atlantique, Nantes Université, École Centrale Nantes, CNRS, INRIA, LS2N, UMR 6004, F-44000 Nantes, France

[†]University of Utah
Salt Lake City, UT, USA

[‡]University of California, San Diego
La Jolla, CA, USA

corresponding contact: daniel.balouek-thomert@inria.fr

Keynote by Manish Parashar

Abstract—The emergence of large-scale cyberinfrastructure composed of heterogeneous computing capabilities and diverse sensors and other data sources are enabling new classes of dynamic data-driven “urgent” applications. However, as the variety of data sources, and the volume and velocity of data grow, processing this data while considering the uncertainty of infrastructure and timeliness constraints of urgent application workflows can be nontrivial and presents a new set of challenges. In this paper, we use an application workflow that monitors and manage the air quality impacts of remote wildfires to illustrate how the R-Pulsar programming system, leveraging the SAGE and WIFIRE platforms, can enable urgent analytics across the computing continuum. R-Pulsar supports urgent data-processing pipelines that tradeoff the content of data, cost of computation, and urgency of the results to support such workflows. We also discuss research challenges associated with programming urgent application workflows and managing resources in an autonomic manner.

Index Terms—Urgent computing, Continuum computing, Edge-Cloud integration, Cyberinfrastructure, Uncertainty

I. INTRODUCTION

Widely-distributed (and pervasive) computing resources and data sources along with low latency connections are enabling new classes of distributed, driven applications, such as those involving urgent analysis of data streams to identify critical events, accelerate decision making, and trigger responses. For example, recent work on Earthquake Early Warning (EEW) builds on such a geographically distributed cyberinfrastructure to integrate data from different types of sensors to increase the accuracy of detection, with efficient and timely computations using a range of resources available along the data path, all while ensuring robustness to data uncertainties and partial infrastructure failures [1], [2].

Indeed, such a *Computing Continuum*, i.e., the spectrum of connected resources and services with heterogeneous capabilities, capacities, availabilities distributed across the logical

extreme of the network and along the data path [3], [4], provides an ideal platform for supporting urgent computing, i.e., computing under strict time and quality constraints to support decision making with the desired confidence within a defined time interval [5]–[7]. However, exploiting resources across this continuum can be challenging [8].

Developing urgent computing applications requires developers and service providers to orchestrate data-driven workflows while considering not only the extreme heterogeneity of underlying cyberinfrastructure, but also the uncertainty related to the availability and quality of the data [9]. Thus, the end-to-end performance of such applications deployed across the continuum depends on programming abstractions to express application behaviors that can react to unforeseen events at runtime, and autonomic runtime mechanisms that support the adaptations of resources and execution paths between the across the continuum. Furthermore, the inherent resource and system dynamics further complicate the realization of urgent computing applications [10]. For example, application developers cannot make assumptions about the exact execution context, which is only know at runtime. As a result, in order to preserve a level of generality, it is essential the programming and runtime systems maintain a separation between the concerns of domain experts (acquainted with the applications) and those of resource managers (familiar with the infrastructure). Furthermore, when choosing resources or carrying out actuation, it is important to take the shared nature of the sensors and resources into consideration.

This paper explores these issues using an application workflow that monitors and manage the air quality impacts of remote wildfires as a driving usecase. It then illustrates how the programming abstractions and runtime mechanisms provided by the R-Pulsar programming system, along with services provided by the SAGE and WIFIRE platforms can be lever-



Fig. 1: Monitoring and managing the air quality impact of wildfires – This driving urgent application usecase is designed using R-Pulsar to composed sensors provided by the SAGE platform align, wildfire data and model provided by the WIFIRE platform along with local pollution concentration maps, all distributed across the computing continuum.

aged potentially enable urgent computing across the computing continuum to support this usecase.

This paper is organized as follows. The driving urgent application usecase scenario is presented in Section II. An overview of the R-Pulsar programming system as well as the SAGE and WIFIRE platforms are described in Section III. Section IV explores how R-Pulsar along with SAGE and WIFIRE can support the urgent application usecase across the continuum and discusses research challenges.

II. MOTIVATING URGENT USECASE: MONITORING AND MANAGING THE AIR QUALITY IMPACTS OF WILDFIRES

High-fidelity data streaming from multiple data sources, combined with highly-accurate computational models, can be effectively used to monitor and manage the impacts of both natural disasters and extreme events. This is illustrated in the wildfire scenario below.

Wildfires are among the most destructive natural disasters, destroying homes, wildlife habitat and timber, and the emissions pollute the air, which is harmful to human health. Among many others in recent times, California’s Dixie Fire burned nearly a thousand square miles in August 2021 [11]. The claims of increased air pollution in Utah as a direct effect of the wildfires in California served as the driver for our research [12]. The motivating scenario is predicated on the utilization of a large number of applied models for the purpose of following and responding to fire events across a variety of states and different data sources. These plugins track wildfire events, track smoke plumes, and model downwind air pollution concentrations associated with wildfire smoke.

These wildfire events might be recognized, predicted, and linked to better alerts and reactions using sensors spread over California, Oregon, and Utah¹. Through a three-step approach, we aim to realize this possibility, illustrated in Figure 1. First, the HPWREN network’s field cameras² are situated in California as a collaborative, networked cyberinfrastructure. To extract photos and increase trust in detecting smoke events,

camera feeds are processed using smoke detection models. Second, a Fire Simulator (FS) gathers latitude/longitude data from the cameras and produces fire metrics relating to the intensity of a starting fire, as well as providing fire emissions as an output. Finally, the Air Quality (AQ) Model runs a Lagrangian transport model with inputs from gridded weather datasets and fire emissions from FS to produce a map of air pollution concentrations. The output is subsequently applied to decision-making or to initiate further sensing. Such a scenario’s specification calls for expertise from several fields and interdisciplinary cooperation.

III. HARNESSING THE COMPUTING CONTINUUM FOR URGENT APPLICATIONS: BUILDING BLOCKS

A. The SAGE Platform

The SAGE cyberinfrastructure platform offers large-scale infrastructure at the national level to enable artificial intelligence (AI) at the network edge. SAGE addresses the tremendous variability of system and service capacities and capabilities³. Extreme uncertainty regarding the availability and quality of data, resources, and services throughout the Edge/Cloud-HPC continuum adds to this heterogeneity. To address this heterogeneity, developers need application formulations that expose natural flexible configurations and scheduling/resource trade-offs as well as establish policies and mechanisms that might trigger runtime adaptations are necessary to address this heterogeneity - the SAGE platform enables the execution of such formulations at the edge [4].

The SAGE catalog of plugins offers applied models that can be deployed on any of their nodes for sensor-based processing such as motion detection using cameras or air quality using dust monitors. These plugins are currently developed regardless of the device or runtime scheduling context in which they are used.

B. The WIFIRE Project

The NSF-funded WIFIRE project [13] develops an integrated system for wildfire monitoring, simulation, and re-

¹<https://www.anl.gov/article/new-sensing-platform-deployed-at-controlled-burn-site-could-help-prevent-forest-fires>

²<http://hpwren.ucsd.edu/cameras/>

³SAGE - Cyber Infrastructure for AI at the Edge - <https://sagecontinuum.org/>

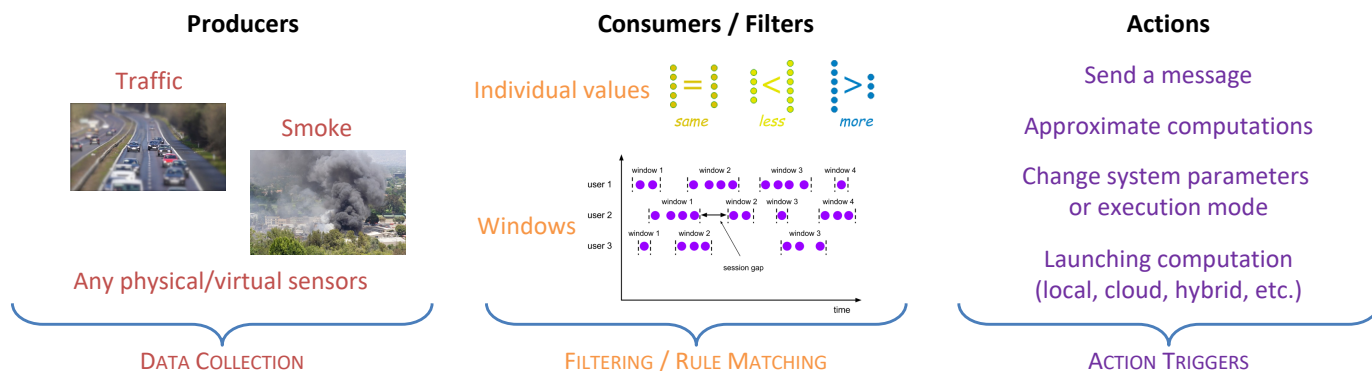


Fig. 2: The R-Pulsar programming system provides content- and location-based abstractions for specifying (1) what data are collected and where the data are analyzed, (2) when to trigger data-processing tasks based on data patterns (spatial and temporal), and (3) what and where to invoke actions as a result of the triggers.

sponse. WIFIRE provides an end-to-end management cyber-infrastructure (CI) with integrated data collection from many real-time data sources, and for modeling wildfire behavior.

C. The R-Pulsar Programming System: Enabling Data-driven Workflows across the Continuum

R-Pulsar is a programming system for building data-to-discovery pipelines between the edge of the network and larger-scale HPC systems and cloud services. It provides middleware services that enables local analytics, messaging, data storage, and data querying capabilities on edge devices using a unified platform [14], [15]. Overall, R-Pulsar offers:

- Content- and location-based programming abstractions for specifying what data are collected and where the data are analyzed.
- Rule-based runtime mechanisms for specifying when and where to trigger data-processing tasks as well as actions based on data patterns (spatial and temporal).

Figure 2 illustrates the key concepts of R-Pulsar. First, producers are in charge of collecting data from physical and virtual sensors. As IoT data often come with temporal and spatial information, R-Pulsar presents a location-aware overlay network to logically connect sensors. This overlay allows developers to specific interactions based on data content and location, rather than using names and/or addresses [16].

Second, R-Pulsar provides an event-driven programming model where rules are triggered using data filters that are based on the availability/content of data and/or the state of resources. Using this model, developers can use filters to specify *IF-THEN* rules to steer the execution of a workflow. In the context of infinite streams of data, the filters can be written to reason on streaming windows to act on specific and/or incomplete batches and may be based on statistical properties of the data in a window. Other types of filters and rules include content-driven, context-aware, or spatial [15].

Finally, actions are triggered when a rule is fired, and can include communications/notifications, further data acquisition/analysis, changing of system parameters or service

configurations, and/or launching additional computations and the implementation of tradeoffs.

We recently deployed and evaluated a policy-driven adaptation of R-Pulsar over the Grid’5000 platform [17] by observing the behavior of the system with variability in data volumes, network bandwidth, and available CPU capacity. Results shows that dynamic policies adjust quickly to periodic variations allowing for the management of a large number of events and complex decisions in cloud-only, edge-only or hybrid setups, with deadline and quality constraints [18].

The latest version of R-Pulsar is available on GitHub⁴ under Apache License 2.0. On-going work is focused on packaging R-Pulsar as lightweight Kubernetes containers for integration with Waggle and SAGE.

IV. IMPLEMENTING URGENT WORKFLOWS ACROSS THE COMPUTING CONTINUUM: APPROACH AND CHALLENGES

A key goal of the research described in this paper is to understand of data-driven feedback necessary to map, schedule, execute and adapt the data-to-discovery workflow from sensors to decision-making, by opportunistically combining heterogeneous and geographically distributed resources. A further goal is to use the presented application usecase to synthesize urgent application patterns underlying similar urgent application workflows, and to develop implementation architectures, based on the R-Pulsar programming system, that can be used to execute these patterns across the edge-to-HPC/cloud computing continuum and realize the desired quality and performance goals.

A. Architecture

In [19], the authors compared and contrasted the state of the art of components and middleware systems that operate between the edge and the cloud, in both academia and industry. The findings from this study highlight a set of common requirements and design goals, for example, for implementing data ingestion, as well as for programming models that allow

⁴<https://http://rpulsar.sci.utah.edu/>

for the processing of streaming data and runtime systems that support the management of resources across the continuum. Figure 3 presents a simplified schematics of a three-layer edge-based middleware architecture for supporting application workflows across the edge-cloud/HPC continuum that emerged from the study.

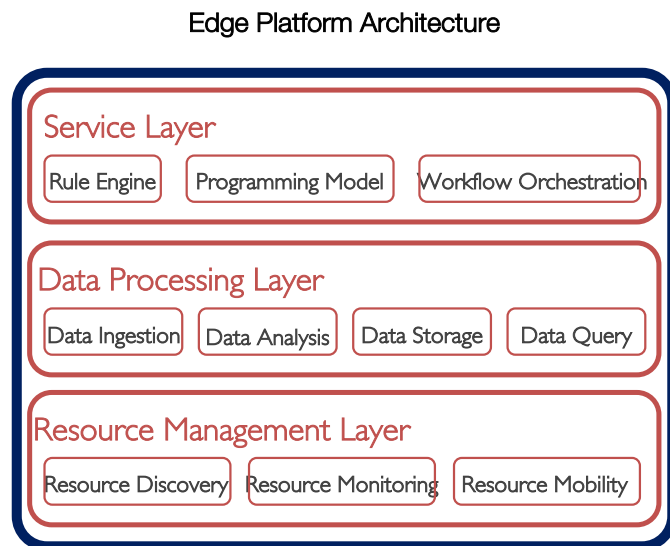


Fig. 3: A schematic overview of an edge-based middleware architecture to support application workflows across the edge-cloud/HPC continuum.

The *Resource Management Layer* organizes the discovery, selection, and allocation of available resources across the heterogeneous and distributed set of available resources. The key design goals of this layer include achieving low overhead, distributed management and aggregating of resources. The *Data Processing Layer* consolidates data from multiple sources, and provides support for data indexing and discovery. The data processing components may need to be able to process data under near-real time constraint and on a range of resource types, including edge constrained devices such as Raspberry Pis or smartphones as well as cloud/HPC systems. The *Service Layer* provides a set of capabilities to support the implementation of application workflows, managing resources, and coordinating execution.

Furthermore, satisfying application requirements and constraints in terms of response time, solution quality, data resolution, cost, energy, etc., during execution across a highly dynamic edge-cloud/HPC continuum warrants optimizations across all three layers. Such cross-layer optimizations must address cost/benefit tradeoffs at runtime in an autonomic manner and may require the reconfiguration of applications, resources, and services, including provisioning new resources and services, adapting scheduling strategies, and/or adjusting application parameters as necessary.

In the context of our driving usecase, i.e., detecting the impacts of wildfire on air quality, the need to reduce execution

time and energy usage requires dynamic resource and service composition and configuration [20].

The on-demand composition of various services (e.g., SAGE plugins) across geographically dispersed resources while considering cross-cutting performance, quality, security or energy concerns is not, however, supported by programming systems or resource management runtimes. For our usecase, sensors are spread across California, Oregon, and Utah, and utilize various analytics plugins for tracking and responding to fire events across these states. The plugins used include, for example, cloud motion, wildfire smoke detection, and air pollution measurements. Composing these distributed sensors and services into data-driven workflows, in spite of the inherent uncertainties (e.g., availability of resources and data, data quality, etc.) remains a challenge – existing workflows require a workflow’s control structure and data-flows to be statically defined. Furthermore, runtime mechanisms must optimize increasingly complex trade-offs between the costs associated with computations, data movement and analysis, etc., and satisfying application requirements for data quality and response times, in a very dynamic execution environment where the availability and status of resources is changing. Ongoing research continues to develop the R-Pulsar programming system and autonomic runtime to address these challenges toward the overarching goal of enabling urgent application workflows across the computing continuum.

V. CONCLUSION

In this paper, we explored how the emerging computing continuum composed of a spectrum of connected resources and services distributed across the logical extreme of the network and along the data path can enable new classes of dynamic data-driven urgent applications. Specifically, we illustrated how the programming abstractions and runtime mechanisms provided by the R-Pulsar programming system, along with services provided by the SAGE and WIFIRE platforms can potentially enable urgent computing across the computing continuum to support an application workflow that monitors and manage the air quality impacts of remote wildfires. We also discussed research challenges associated with programming urgent application workflows and managing resources in an autonomic manner.

ACKNOWLEDGEMENTS

This research is supported in part by the NSF under grant number OAC 2238064, OAC 1640834, OAC 1835692. The research was conducted as part of the SCI Institute at the University of Utah.

REFERENCES

- [1] K. Fauvel, D. Daniel Balouek-Thomert, D. Melgar, P. Silva, A. Simonet, G. Antoniu, A. Costan, V. Masson, M. Parashar, I. Rodero, and A. Termier. A Distributed Multi-Sensor Machine Learning Approach to Earthquake Early Warning. In *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence*, 2020.
- [2] Marco Esposito, Lorenzo Palma, Alberto Belli, Luisiana Sabbatini, and Paola Pierleoni. Recent advances in internet of things solutions for early warning systems: A review. *Sensors*, 22(6):2124, 2022.

- [3] D. Balouek-Thomert, E. G. Renart, A. R. Zamani, A. Simonet, and M. Parashar. Towards a computing continuum: Enabling edge-to-cloud integration for data-driven workflows. *The International Journal of High Performance Computing Applications*, 33(6):1159–1174, November 2019.
- [4] P. Beckman, J. Dongarra, N. Ferrier, G. Fox, T. Moore, D. Reed, and M. Beck. Harnessing the computing continuum for programming our world, 2019.
- [5] Daniel Balouek-Thomert, Eddy Caron, Laurent Lefevre, and Manish Parashar. Towards a methodology for building dynamic urgent applications on continuum computing platforms. In *2022 First Combined International Workshop on Interactive Urgent Supercomputing (CIW-IUS)*, pages 1–6, 2022.
- [6] Siew Hoon Leong and Dieter Kranzlmüller. Towards a general definition of urgent computing. *Procedia Computer Science*, 51:2337 – 2346, 2015. International Conference On Computational Science, ICCS 2015.
- [7] Alexander V Boukhanovsky, Valeria V Krzhizhanovskaya, and Marian Bubak. Urgent computing for decision support in critical situations, 2018.
- [8] Daniel Balouek-Thomert, Ivan Rodero, and Manish Parashar. Harnessing the computing continuum for urgent science. *ACM SIGMETRICS Performance Evaluation Review*, 48(2):41–46, 2020.
- [9] Sergey V. Kovalchuk, Pavel A. Smirnov, Sergey V. Maryin, Timofey N. Tchurov, and Vladislav A. Karbovskiy. Deadline-driven resource management within urgent computing cyberinfrastructure. *Procedia Computer Science*, 18:2203–2212, 2013. 2013 International Conference on Computational Science.
- [10] Gordon Gibb, Rupert Nash, Nick Brown, and Bianca Prodan. The technologies required for fusing hpc and real-time data to support urgent computing. In *2019 IEEE/ACM HPC for Urgent Decision Making (UrgentHPC)*, pages 24–34, 2019.
- [11] Cal Fire, Shasta County, and Unified Command Fires. California statewide fire summary. *Ertugrul, M., Varol, T., Ozel, HB, Cetin, M. and Sevik, H*, pages 1–17, 2021.
- [12] Liji M David, AR Ravishankara, Steven J Brey, Emily V Fischer, John Volckens, and Sonia Kreidenweis. Could the exception become the rule? uncontrollable air pollution events in the us due to wildland fires. *Environmental Research Letters*, 16(3):034029, 2021.
- [13] Ilkay Altintas, Jessica Block, Raymond De Callafon, Daniel Crawl, Charles Cowart, Amarnath Gupta, Mai Nguyen, Hans-Werner Braun, Jurgen Schulze, Michael Gollner, et al. Towards an integrated cyber-infrastructure for scalable data-driven monitoring, dynamic prediction and resilience of wildfires. *Procedia Computer Science*, 51:1633–1642, 2015.
- [14] Eduard Renart, Daniel Balouek-Thomert, Xuan Hu, Jie Gong, and Manish Parashar. Online decision-making using edge resources for content-driven stream processing. In *2017 IEEE 13th International Conference on e-Science (e-Science)*, pages 384–392. IEEE, 2017.
- [15] Eduard Gibert Renart, Daniel Balouek-Thomert, and Manish Parashar. An edge-based framework for enabling data-driven pipelines for iot systems. In *2019 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pages 885–894. IEEE, 2019.
- [16] Nanyan Jiang, Andres Quiroz, Cristina Schmidt, and Manish Parashar. Meteor: a middleware infrastructure for content-based decoupled interactions in pervasive grid environments. *Concurrency and Computation: Practice and Experience*, 20(12):1455–1484, 2008.
- [17] Daniel Balouek, Alexandra Carpen Amarie, Ghislain Charrier, Frédéric Desprez, Emmanuel Jeannot, Emmanuel Jeanvoine, Adrien Lèbre, David Margery, Nicolas Niclausse, Lucas Nussbaum, et al. Adding virtualization capabilities to the grid’5000 testbed. In *International Conference on Cloud Computing and Services Science*, pages 3–20. Springer, Cham, 2012.
- [18] Daniel Balouek-Thomert, Ivan Rodero, and Manish Parashar. Evaluating policy-driven adaptation on the edge-to-cloud continuum. In *2021 IEEE/ACM HPC for Urgent Decision Making (UrgentHPC)*, pages 11–20, 2021.
- [19] Daniel Balouek-Thomert and Manish Parashar. Platforms for edge computing and internet of things applications: A survey. In *2021 Thirteenth International Conference on Contemporary Computing (IC3-2021)*, IC3 ’21, page 140–149, New York, NY, USA, 2021. Association for Computing Machinery.
- [20] Ilkay Altintas, Ismael Perez, Dmitry Mishin, Adrien Trouillaud, Christopher Irving, John Graham, Mahidhar Tatineni, Thomas DeFanti, Shawn Strande, Larry Smarr, and Michael L. Norman. Towards a dynamic composability approach for using heterogeneous systems in remote sensing. In *2022 IEEE 18th International Conference on e-Science (e-Science)*, pages 336–345, 2022.