Accepted: 11 September 2025
Revised: 10 September 2025
Received: 10 September 2025

RESEARCH-ARTICLE

Autonomic Computing Rebooted: Taming the Computing Continuum

MANISH PARASHAR, The University of Utah, Salt Lake City, UT, United States

ACM Transactions on Autonomous and Adaptive Systems
https://doi.org/10.1145/3768320
EISSN: 1556-4703

Autonomic Computing Rebooted: Taming the Computing Continuum

MANISH PARASHAR*, Scientific Computing and Imaging (SCI) Institute, University of Utah, USA

Technological advances and rapid service deployments have resulted in a pervasive and interconnected *computing continuum*, which is enabling new classes of application formulations and workflows, delivering novel services to consumers, and is becoming a core engine for discovery, innovation, and economic growth. The computing continuum is also unleashing new system and application management challenges, which must be addressed before its potential and promises are truly realized. Autonomic computing can provide the abstractions and mechanisms essential to effectively harnessing the computing continuum, but must evolve to address these new challenges. This paper is a call to action for rebooting autonomics to enable us to harness the computing continuum.

CCS Concepts: • **General and reference → Cross-computing tools and techniques**.

1 AN EMERGING COMPUTING CONTINUUM

Technological advances and rapid service deployments have resulted in a pervasive and interconnected *computing continuum* that seamlessly integrates distributed and diverse data sources with nontrivial data processing capabilities at the edge and along the data path, as well as extreme-scale enterprise computing systems and HPC supercomputers. For example, a natural continuum is evolving across the science ecosystem, spanning large-scale instruments, experimental facilities, observatories, and sensor networks, all streaming data; high-speed networks and network services; and a range of computing capabilities along the continuum, from edge to in-network to data center/HPC.

This computing continuum, which is everywhere and nowhere [5], is enabling new classes of application formulations and workflows, delivering novel services to consumers, and is becoming a core engine for discovery, innovation, and economic growth. In science, such application workflows combine sensing and streaming data (e.g., from sensor networks, observatories, or experimental facilities) with simulations and data-driven modeling, and actuation to understand, analyze, predict, and actuate.

One class of scientific applications workflow that is enabled by the continuum and is being increasingly deployed is end-to-end experiment management, where streaming data from an experiment or instrument is analyzed and modeled, and the result of the modeling is used to control, manage, and/or optimize the experiment. Another growing class of application workflows is digital twins for large-scale complex systems. These are digital representations of actual, real-world physical systems and can serve as vehicles for understanding, managing, optimizing, and protecting the physical systems [4]. These application workflows highlight the need for combining real-time data acquisition with large-scale modeling, both data-driven and mathematical, and possible actuation. The computing continuum will play a significant role in making these systems a reality.

*This is an invited reflection in celebration of ACM TAAS 20th anniversary.

Author's Contact Information: Manish Parashar, manish.parashar@utah.edu, Scientific Computing and Imaging (SCI) Institute, University of Utah, Salt Lake City, Utah, USA.

ACM 1556-4703/2025/9-ART
https://doi.org/10.1145/3768320

An important class of applications enabled by the computing continuum is urgent science. Urgent science refers to science that can support such rigorous, data-driven decision-making in response to emergencies, while effectively managing inherent uncertainties. Urgent computing supports urgent science. It can be defined as computing under strict time and quality constraints to support decision-making with the desired confidence, within a specified time interval [2, 3]. The goal is to leverage data and computations to support decision-making during an emergency. Urgent computing workflows utilize the computing continuum to access data from various sources, along with other resources and services, to detect events, develop a response, and trigger actions.

2 REBOOTING AUTONOMICS FOR THE COMPUTING CONTINUUM

The computing continuum is unleashing new system and application management challenges, which must be addressed before its potential and promises are truly realized. Autonomic computing [7, 8] emerged in the early 2000s as an effective approach for addressing similar challenges within enterprise datacenters. Its overarching goal is to realize computer and software systems and applications that can manage themselves in accordance with high-level guidance from humans. It provides an alternative paradigm for system and application design, inspired by strategies that biological systems have evolved to address similar challenges.

Autonomic computing can provide the abstractions and mechanisms essential to effectively harnessing the computing continuum by introducing self-management capabilities to address its complexity, heterogeneity, and dynamic nature. The computing continuum introduces new levels of distribution, heterogeneity, dynamic behaviors, uncertainties, vulnerabilities, and overall operational complexity, and consequently, autonomic management techniques have to evolve to address these complexities. At the same time, recent advances in artificial intelligence (AI), including agentic AI, provide new opportunities for autonomic management – it is transforming automated management from a static rules-bound process into an intelligent, adaptive one. By embedding machine learning into autonomic control loops, systems can learn from operational data to improve operations (e.g., scheduling, fault tolerance, and energy efficiency) and application workflow execution across the computing continuum. As a result, it is time to reboot autonomic computing and evolve along multiple dimensions, including:

- *Wide Distribution/Edge and Cloud Integration:* Autonomics has to move beyond the datacenter to a widely distributed and highly dynamic edge+cloud/HPC continuum.
- *AI-Augmented Autonomics:* The infusion of Large Language Models, Agentic AI frameworks, and other AI solutions and AI-enabled mechanisms into system management is opening new possibilities for self-optimization and resilience.
- *Resilience and Trust:* Managing uncertainty, quantifying error, and ensuring system stability remain central challenges—especially in mission-critical and scientific computing contexts.

3 A COMPELLING RESEARCH AGENDA

While autonomic computing provides a promising approach for addressing the many underlying research challenges in making urgent computing and data-driven workflows executing across the computing continuum a reality, it does present a compelling research agenda. Furthermore, complementing the fundamental advances needed to address these research questions, translational research is equally critical to achieve the urgent computing vision [1]. Translational computer science research aims to closely couple cycles of innovation between computer science and other disciplines to accelerate and amplify its transformative impact.

For example, in urgent computing, autonomic resource provisioning and scheduling decisions would be based on strict time constraints, data and resource availability, data quality, acceptable uncertainty, etc. More importantly, rather than typical autonomic optimization objectives, autonomic behaviors for these application workflows would focus on cost/benefit tradeoffs, determining what is “good enough,” to support decision-making.

Note that AI-driven autonomics inherits many of the challenges faced by AI itself, such as security, privacy and trust, explainability, and accountability, all of which are extremely important for autonomic systems.

As noted earlier, data-driven application workflows enabled by the computing continuum often involve real-time data streaming from the edge and transformed along its data path, scalable analytics in the cloud, and intensive simulations on HPC platforms. Autonomic systems—those capable of self-configuration, self-healing, self-optimization, and self-protection—can orchestrate these diverse components efficiently and adaptively to meet application objectives and system efficiencies and qualities for service. For example, autonomic resource management can dynamically allocate workloads across edge, cloud, and HPC layers based on performance, energy, and latency constraints, autonomic orchestration can ensure seamless workflow composition and execution across heterogeneous infrastructures and services, and autonomic workflow execution can use user-defined policies along with system constraints to implement necessary tradeoffs at runtime. Such autonomic services are particularly valuable for use cases such as digital twins, smart cities, and precision agriculture, where data flows and computational demands vary continuously and unpredictably. The National Data Platform [6] [1], with an overarching goal of democratizing data access and use across the computing continuum, is one project that aims to address many of these challenges and enable data-driven, AI-enabled workflows underlying these applications.

REFERENCES

[1] David Abramson and Manish Parashar. 2019. Translational Research in Computer Science. *Computer* 52, 9 (2019), 16–23. doi:10.1109/MC.2019.2925650

[2] Daniel Balouek-Thomert, Eduard Gibert Renart, Ali Reza Zamani, Anthony Simonet, and Manish Parashar. 2019. Towards a Computing Continuum: Enabling Edge-to-Cloud Integration for Data-Driven Workflows. *The International Journal of High Performance Computing Applications* 33, 6 (2019), 1159–1174. arXiv:https://doi.org/10.1177/1094342019877383 doi:10.1177/1094342019877383

[3] Daniel Balouek-Thomert, Ivan Rodero, and Manish Parashar. 2020. Harnessing the Computing Continuum for Urgent Science. *SIGMETRICS Perform. Eval. Rev.* 48, 2 (Nov. 2020), 41–46. doi:10.1145/3439602.3439618

[4] National Academy of Engineering, Engineering National Academies of Sciences, and Medicine. 2024. *Foundational Research Gaps and Future Directions for Digital Twins*. The National Academies Press, Washington, DC. doi:10.17226/26894

[5] Manish Parashar. 2025. Everywhere and Nowhere: Envisioning a Computing Continuum for Science. *Computing in Science & Engineering* 27, 1 (2025), 51–56. doi:10.1109/MCSE.2025.3543924

[6] Manish Parashar and Ilkay Altintas. 2023. Toward Democratizing Access to Science Data: Introducing the National Data Platform. In *2023 IEEE 19th International Conference on e-Science (e-Science)*. 1–4. doi:10.1109/e-Science58273.2023.10254930

[7] Manish Parashar and Salim Hariri. 2005. Autonomic Computing: An Overview. In *Unconventional Programming Paradigms*, Jean-Pierre Banâtre, Pascal Fradet, Jean-Louis Giavitto, and Olivier Michel (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 257–269.

[8] M. Parashar and S. Hariri. 2018. *Autonomic Computing: Concepts, Infrastructure, and Applications*. CRC Press. https://books.google.com/books?id=OgJ-DwAAQBAJ

[1] The National Data Platform (NDP) Pilot Project, https://nationaldataplatform.org/.