



HAL
open science

Toward AI-based Complex Self-Adaptive Systems

Brell Peclard Sanwouo, Clément Quinton, Paul Temple

► **To cite this version:**

Brell Peclard Sanwouo, Clément Quinton, Paul Temple. Toward AI-based Complex Self-Adaptive Systems. BENEVOL'24 - 23rd Belgium-Netherlands Software Evolution Workshop, Nov 2024, Namur, Belgium. hal-04896457v1

HAL Id: hal-04896457

<https://inria.hal.science/hal-04896457v1>

Submitted on 20 Jan 2025 (v1), last revised 20 Jan 2025 (v2)

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

Toward a AI-based complex Self-Adaptive Systems.

Brell P. SANWOUO¹, Clément Quinton¹ and Paul Temple²

¹Univ. Lille, CNRS, Inria, UMR 9189 CRISTAL, F-59000 Lille, France

²Univ. Rennes, CNRS, Inria, IRISA, Rennes, France

Abstract

Cloud computing has become an essential pillar of modern IT infrastructures, offering storage and computing capacity on an unprecedented scale. Using this technology, businesses can benefit from great flexibility, almost infinite scalability, and global accessibility, while significantly reducing infrastructure costs. In parallel, the emergence of microservice architectures [1, 2] is revolutionizing the implementation of large-scale software infrastructures and architectures. Although microservice architectures offer modularity, scalability and flexibility, they also pose challenges in deployment, management, latency optimization, and testing. It is crucial to implement strategies to understand user intentions, simplify tasks, detect issues proactively, analyze error causes and maintain system reliability.

To achieve this, each microservice is closely monitored and corrective actions are regularly implemented. This process can be framed by the MAPE-K loop (Monitor, Analyze, Plan, Execute, Knowledge) proposed by Kephart et al. (2003) [3], which provides a methodological framework for autonomous service management. By monitoring services, analyzing the collected data, planning actions, executing these plans and using the knowledge acquired to continually improve the process, the MAPE-K loop helps to maintain the alignment and coherence of the microservice architecture.

Self-adaptive systems (SAS) are engineered to modify their operations dynamically in response to environmental variations. The integration of AI, especially machine learning, into these systems not only facilitates real-time responses to such changes but also helps in predicting and meeting future requirements through optimized solutions. Previous studies have highlighted the significant enhancements in adaptability when machine learning models are incorporated into the MAPE-K feedback loop of these systems.

For example, Araújo et al. (2021) [4] have developed a tool called Finch, which integrates machine learning into the MAPE-K loop to automate the configuration and optimisation of REST systems, thus reducing the manual efforts. Similarly, Weyns et al. (2021) [5] explored the combination of MAPE-K approaches, control theory principles, and machine learning techniques, demonstrating that this integration can produce more robust and efficient adaptive systems. Ouareth et al. (2022) [6] used reinforcement learning to enable dynamic adjustments in the structure and behaviour of MAPE-K loops, which is particularly effective for complex systems requiring decentralised coordination. Lecheheb et al. (2022) [7] have proposed a hybrid approach combining machine learning and Deep Learning to improve the accuracy of the Analyzer component in the MAPE-K loop, thereby reducing the time and resources required to adapt large-scale systems.

Nowadays, with the rise of Generative Artificial Intelligence (GenAI), which in certain ways offers a better alternative to machine learning, research into self-adaptive systems has continued. For example, Nakagawa et al. (2023) [8] explore the use of GenAI to automate goal model generation based on the use of the MAPE-K adaptation loop. This approach aims to improve the efficiency and accuracy of self-adaptive systems by exploiting the capabilities of GenAI to create more robust and context-sensitive adaptation strategies. More specifically, Donakanti et al. (2024) [9] study how LLMs can improve software system adaptation at the architecture level by autonomously generating context-sensitive adaptation strategies, just as humans would. On the other hand, Li et al. (2024) [10, 11] discuss the specific benefits and challenges of employing GenAI, and in particular Large Language Models (LLMs), in self-adaptive systems. Despite the technological diversity and rapid evolution of GenAI, these papers highlight that there are still gaps in understanding the practical benefits of integrating these technologies into self-adaptive systems. Similarly, Nunes et al. (2024) [12] propose MS-RA, a self-adaptive, requirements-driven solution for microservices autoscaling. This approach adjusts resource consumption to meet changing workload demands, ensuring system performance and cost-effectiveness while using fewer resources.

As several directions emerge with the use of GenAI in the context of self-adaptive systems, we would like to focus on improving GenAI and in particular LLMs, by exploring advanced techniques such as diffusion models, Mixture of Experts, and RLHF (Reinforcement Learned via Human Feedback). These methods aim to increase the efficiency of LLMs in complex contexts, enabling them to interact more effectively with dynamic data in real time, particularly during the monitoring and analysis phases. We will also study how the optimisation of the prompts used by GenAI can be aligned with the objectives of each stage of the MAPE-K loop. By refining the prompts for each phase, we hope not only to improve anomaly detection during monitoring, but also to refine the analysis capabilities of GenAI to support all stages of adaptation. Finally, we would like to analyze how different ways of structuring and visualizing data, particularly in the monitoring phase, can facilitate the integration of GenAI into the self-adaptation process, while improving system capacity.

Keywords

Large Languages Models, Generative AI, Artificial Intelligence, Machine Learning, Self-Adaptative Systems, Autonomous Systems

References

- [1] M. Viggiato, R. Terra, H. Rocha, M. T. Valente, E. Figueiredo, Microservices in practice: A survey study, arXiv preprint arXiv:1808.04836 (2018).
- [2] N. Dragoni, S. Giallorenzo, A. L. Lafuente, M. Mazzara, F. Montesi, R. Mustafin, L. Safina, Microservices: yesterday, today, and tomorrow, *Present and ulterior software engineering (2017)* 195–216.
- [3] J. O. Kephart, D. M. Chess, The vision of autonomic computing, *Computer* 36 (2003) 41–50.
- [4] R. Araújo, R. Holmes, Lightweight self-adaptive configuration using machine learning, in: *Proceedings of the 31st Annual International Conference on Computer Science and Software Engineering*, 2021, pp. 133–142.
- [5] D. Weyns, B. Schmerl, M. Kishida, A. Leva, M. Litoiu, N. Ozay, C. Paterson, K. Tei, Towards better adaptive systems by combining mape, control theory, and machine learning, in: *2021 International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS)*, IEEE, 2021, pp. 217–223.
- [6] S. Ouareth, S. Boulehouache, M. Smaine, Self-adaptation through reinforcement learning using a feature model, *International Journal of Organizational and Collective Intelligence (IJOICI)* 12 (2022) 1–20.
- [7] S. Lecheheb, S. Boulehouache, S. Brahimi, Improving self-adaptation by combining mape-k, machine and deep learning, in: *2022 2nd International Conference on New Technologies of Information and Communication (NTIC)*, IEEE, 2022, pp. 1–6.
- [8] H. Nakagawa, S. Honiden, Mape-k loop-based goal model generation using generative ai, in: *2023 IEEE 31st International Requirements Engineering Conference Workshops (REW)*, 2023, pp. 247–251. doi:10.1109/REW57809.2023.00050.
- [9] R. Donakanti, P. Jain, S. Kulkarni, K. Vaidhyanathan, Reimagining self-adaptation in the age of large language models, arXiv preprint arXiv:2404.09866 (2024).
- [10] J. Li, M. Zhang, N. Li, D. Weyns, Z. Jin, K. Tei, Generative ai for self-adaptive systems: State of the art and research roadmap, *ACM Trans. Auton. Adapt. Syst.* (2024). URL: <https://doi.org/10.1145/3686803>. doi:10.1145/3686803.
- [11] J. Li, M. Zhang, N. Li, D. Weyns, Z. Jin, K. Tei, Exploring the potential of large language models in self-adaptive systems, in: *Proceedings of the 19th International Symposium on Software Engineering for Adaptive and Self-Managing Systems*, 2024, pp. 77–83.
- [12] J. P. Karol Santos Nunes, S. Nejati, M. Sabetzadeh, E. Y. Nakagawa, Self-adaptive, requirements-driven autoscaling of microservices, in: *Proceedings of the 19th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, SEAMS '24*, Association for Computing Machinery, New York, NY, USA, 2024, p. 168–174. URL: <https://doi.org/10.1145/3643915.3644094>. doi:10.1145/3643915.3644094.

BENEVOL: Belgium-Netherlands Software Evolution Workshop, November 21–22, 2024, Namur, Belgium

✉ brell-peclard.sanwouo-chekam@inria.fr (B. P. SANWOUO); clement.quinton@inria.fr (C. Quinton); paul.temple@irisa.fr (P. Temple)

ORCID 0000-0003-3203-6107 (C. Quinton); 0000-0002-8276-0593 (P. Temple)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).