DAFTAR PUSTAKA

Al Qassem, L. M., Stouraitis, T., Damiani, E. and Elfadel, I. A. M. (2023). Proactive Random-Forest Autoscaler for Microservice Resource Allocation. *IEEE Access*, 11: 2570–2585.

Bartelucci, N., Bellavista, P., Pusztai, T., Morichetta, A. and Dustdar, S. (2022). High-Level Metrics for Service Level Objective-aware Autoscaling in Polaris: a Performance Evaluation. In: *2022 IEEE 6th International Conference on Fog and Edge Computing (ICFEC)*. pp.73–77.

Bi, T., Pan, Y., Jiang, X., Ma, M. and Wang, P. (2022). VECROsim: A Versatile Metric-oriented Microservice Fault Simulation System (Tools and Artifact Track). In: *2022 IEEE 33rd International Symposium on Software Reliability Engineering (ISSRE)*. pp.297–308.

Cheng, K., Zhang, S., Tu, C., Shi, X., Yin, Z., Lu, S., Liang, Y. and Gu, Q. (2023). ProScale: Proactive Autoscaling for Microservice With Time-Varying Workload at the Edge. *IEEE Transactions on Parallel and Distributed Systems*, 34: 1294–1312.

Choi, B., Park, J., Lee, C. and Han, D. (2021). pHPA: A Proactive Autoscaling Framework for Microservice Chain. In: *5th Asia-Pacific Workshop on Networking (APNet 2021)*. Shenzhen, China China: ACM. pp.65–71. https://dl.acm.org/doi/10.1145/3469393.3469401. Accessed 17 October 2022.

Chouliaras, S. and Sotiriadis, S. (2022). Auto-scaling containerized cloud applications: A workload-driven approach. *Simulation Modelling Practice and Theory*, 121: 102654.

Dang-Quang, N.-M. and Yoo, M. (2022). An Efficient Multivariate Autoscaling Framework Using Bi-LSTM for Cloud Computing. *Applied Sciences*, 12: 3523.

Dang-Quang, N.-M. and Yoo, M. (2021). Deep Learning-Based Autoscaling Using Bidirectional Long Short-Term Memory for Kubernetes. *Applied Sciences*, 11: 3835.

Garriga, M. (2018). Towards a Taxonomy of Microservices Architectures. In: A. Cerone & M. Roveri (eds) *Software Engineering and Formal Methods*. Cham: Springer International Publishing. pp.203–218.

He, Z. (2020). Novel Container Cloud Elastic Scaling Strategy based on Kubernetes. In: *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*. Chongqing, China: IEEE. pp.1400–1404. https://ieeexplore.ieee.org/document/9141552/. Accessed 17 October 2022.

Herbst, N., Krebs, R., Oikonomou, G., Kousiouris, G., Evangelinou, A., Iosup, A. and Kounev, S. (2016). Ready for Rain? A View from SPEC Research on the Future of Cloud Metrics. http://arxiv.org/abs/1604.03470. Accessed 21 July 2023.

Horn, A., Fard, H. M. and Wolf, F. (2022). Multi-objective Hybrid Autoscaling of Microservices in Kubernetes Clusters. In: *Euro-Par 2022: Parallel*

Processing. Springer, Cham. pp.233–250. https://link.springer.com/chapter/10.1007/978-3-031-12597-3_15. Accessed 26 January 2023.

Horovitz, S. and Arian, Y. (2018). Efficient Cloud Auto-Scaling with SLA Objective Using Q-Learning. In: *2018 IEEE 6th International Conference on Future Internet of Things and Cloud (FiCloud)*. pp.85–92.

Imdoukh, M., Ahmad, I. and Alfailakawi, M. Gh. (2020a). Machine learning-based auto-scaling for containerized applications. *Neural Computing and Applications*, 32: 9745–9760.

Imdoukh, M., Ahmad, I. and Alfailakawi, M. Gh. (2020b). Machine learning-based auto-scaling for containerized applications. *Neural Computing and Applications*, 32: 9745–9760.

Khaleq, A. A. and Ra, I. (2021). Intelligent Autoscaling of Microservices in the Cloud for Real-Time Applications. *IEEE Access*, 9: 35464–35476.

Kubernetes Authors *Kubernetes Documentation* (on-line). https://kubernetes.io/docs/home/. Accessed 17 October 2022.

Meng, C., Tong, J., Pan, M. and Yu, Y. (2022). HRA: An Intelligent Holistic Resource Autoscaling Framework for Multi-service Applications. In: *2022 IEEE International Conference on Web Services (ICWS)*. pp.129–139.

Nashold, L. and Krishnan, R. (2020). Using LSTM and SARIMA Models to Forecast Cluster CPU Usage. http://arxiv.org/abs/2007.08092. Accessed 17 October 2022.

Qiu, H., Banerjee, S. S., Jha, S., Kalbarczyk, Z. T. and Iyer, R. K. (2020). FIRM: an intelligent fine-grained resource management framework for SLO-oriented microservices. In: *Proceedings of the 14th USENIX Conference on Operating Systems Design and Implementation*. USA: USENIX Association. pp.805–825.

Sampaio, L., Goes, A., Albuquerque, M., Gama, D., Schmid, J. I. and Brito, A. (2020). Single-Input Multiple-Output Control for Multi-Goal Orchestration. In: *2020 IEEE/ACM 13th International Conference on Utility and Cloud Computing (UCC)*. pp.206–215.

Sfakianakis, Y., Marazakis, M. and Bilas, A. (2021). Skynet: Performance-driven Resource Management for Dynamic Workloads. In: *2021 IEEE 14th International Conference on Cloud Computing (CLOUD)*. pp.527–539.

Toka, L., Dobreff, G., Fodor, B. and Sonkoly, B. (2020). Adaptive AI-based auto-scaling for Kubernetes. In: *2020 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID)*. Melbourne, Australia: IEEE. pp.599–608. https://ieeexplore.ieee.org/document/9139654/. Accessed 17 October 2022.

Tran, M.-N., Vu, D.-D. and Kim, Y. (2022). A Survey of Autoscaling in Kubernetes. In: *2022 Thirteenth International Conference on Ubiquitous and Future Networks (ICUFN)*. pp.263–265.

Wajahat, M., Karve, A., Kochut, A. and Gandhi, A. (2019). MLscale: A machine learning based application-agnostic autoscaler. *Sustainable Computing: Informatics and Systems*, 22: 287–299.

Wang, C., Yoshikane, N. and Tsuritani, T. (2021). Usage of a Graph Neural Network for Large-Scale Network Performance Evaluation. In: *2021 International Conference on Optical Network Design and Modeling (ONDM)*. pp.1–5.

Yan, M., Liang, X., Lu, Z., Wu, J. and Zhang, W. (2021). HANSEL: Adaptive horizontal scaling of microservices using Bi-LSTM. *Applied Soft Computing*, 105: 107216.

Yu, B., Yin, H. and Zhu, Z. (2018). Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. : 3634–3640.

Zhang, F., Tang, X., Li, X., Khan, S. U. and Li, Z. (2019). Quantifying cloud elasticity with container-based autoscaling. *Future Generation Computer Systems*, 98: 672–681.