Budget and Performance-efficient Application Deployment along Edge-Fog-Cloud Ecosystem

Polona Štefanič
Cardiff University, School of Computer Science & Informatics
Queen's Buildings, 5 The Parade, CF243AA, UK
Email: StefanicP@cardiff.ac.uk

Omer F. Rana
Cardiff University, School of Computer Science & Informatics
Queen's Buildings, 5 The Parade, CF243AA, UK
Email: RanaOF@cardiff.ac.uk

Vlado Stankovski
University of Ljubljana, Faculty of Civil and Geodetic Engineering
Jamova cesta 2, 1000 Ljubljana, Slovenia
Email: vlado.stankovski@fgg.uni-lj.si

ABSTRACT

Applications that make use of Internet of Things (IoT) capture an enormous amount of raw data from sensors and actuators, which is frequently transmitted towards the cloud data centres for processing and analysis. However, due to varying and unpredictable data generation rates and network latency, sending the data towards a cloud data centre can lead to a performance bottleneck. With the emergence of Fog and Edge computing hosted microservices, data processing could be moved towards the network edge. We propose a novel Pareto-based approach that makes use of a multi-criteria bin packing optimisation for efficient and optimal distributed deployment of microservices – along edge, fog/cloudlet and cloud tiers. This optimisation takes account of non-functional requirements, such as operational cost, compute resource utilisation, service availability, response time, latency and similar. The results show that the present approach provides an optimal and sustainable consumption of compute resources and improves Quality of Service of the application during its runtime. The approach can also be integrated into software engineering workbenches for the creation and deployment of cloud-native applications, enabling partitioning of an application across the multiple infrastructure tiers outlined above.

Keywords—Edge/Fog/Cloud computing, Quality of Service, IoT, Microservice, Pareto front

REFERENCES

[1] G. Burrows, "Harnessing the Internet of Things for business benefit," *The Nation*, 2015. [Online]. Available: https://images.forbes.com/forbesinsights/pitney bowes iot/HarnessingTheInternetofThings.pdf

[2] D. Evans, "The Internet of Things - How the Next Evolution of the Internet is Changing Everything," *CISCO white paper*, no. April, pp. 1–11, 2011. [Online]. Available: http://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:The+Internet+of+Things+-+How+the+Next+Evolution+of+the+Internet+is+Changing+Everything{#}0

[3] P. Mell and T. Grance, "The NIST Definition of Cloud Computing Recommendations of the National Institute of Standards and Technology," *National Institute of Standards and Technology, Information Technology Laboratory*, vol. 145, p. 7, 2011. [Online]. Available: http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

[4] Y. Simmhan, S. Aman, A. Kumbhare, R. Liu, S. Stevens, Q. Zhou, and V. Prasanna, "Cloud-based software platform for big data analytics in smart grids," *Computing in Science Engineering*, vol. 15, no. 4, pp. 38–47, July 2013.

[5] P. A. Legg, D. H. S. Chung, M. L. Parry, R. Bown, M. W. Jones, I. W. Griffiths, and M. Chen, "Transformation of an uncertain video search pipeline to a sketch-based visual analytics loop," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2109–2118, Dec 2013.

[6] Q. Pu, G. Ananthanarayanan, P. Bodik, S. Kandula, A. Akella, P. Bahl, and I. Stoica, "Low latency geo-distributed data analytics," in *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*, ser. SIGCOMM '15. New York, NY, USA: ACM, 2015, pp. 421–434. [Online]. Available: http://doi.acm.org/10.1145/2785956.2787505

[7] M. Villari, M. Fazio, S. Dustdar, O. Rana, and R. Ranjan, "Osmotic computing: A new paradigm for edge/cloud integration," *IEEE Cloud Computing*, vol. 3, no. 6, pp. 76–83, Nov 2016.

[8] F. Bonomi, R. A. Milito, P. Natarajan, and J. Zhu, "Fog computing: A platform for internet of things and analytics," in *Big Data and Internet of Things*, 2014.

[9] Y. Wu, C. Wu, B. Li, X. Qiu, and F. C. M. Lau, "Cloudmedia: When cloud on demand meets video on demand," in *2011 31st International Conference on Distributed Computing Systems*, June 2011, pp. 268–277.

[10] A. R. Zamani, M. Zou, J. Diaz-Montes, I. Petri, O. Rana, A. Anjum, and M. Parashar, "Deadline constrained video analysis via in-transit computational environments," *IEEE Transactions on Services Computing*, vol. PP, pp. 1–1, 01 2017.

[11] G. Ananthanarayanan, P. Bahl, P. Bod´ık, K. Chintalapudi, M. Philipose, L. Ravindranath, and S. Sinha, "Real-time video analytics: The killer app for edge computing," *Computer*, vol. 50, no. 10, pp. 58–67, 2017.

[12] L. Knight, P. Stefanic, M. Cigale, A. C. Jones, and I. J. Taylor, "Towards a methodology for creating time-critical, cloud-based CUDA applications," in *Proceedings of IT4RIs 18: Interoperable infrastructures for interdisciplinary big data sciences- Time critical applications and infrastructure optimization*, Amsterdam, 2018. [Online]. Available: https://doi.org/10.5281/zenodo.1162877

[13] O. Rana, M. Shaikh, M. Ali, A. Anjum, and L. Bittencourt, "Vertical workflows: Service orchestration across cloud amp; edge resources," in *2018 IEEE 6th International Conference on Future Internet of Things and Cloud (FiCloud)*, Aug 2018, pp. 355–362.

[14] M. Kassab, M. Daneva, and O. Ormandjieva, "Scope management of non-functional requirements," in *Proceedings of the 33rd EUROMICRO Conference on Software Engineering and Advanced Applications*, ser. EUROMICRO '07. Washington, DC, USA: IEEE Computer Society, 2007, pp. 409–417. [Online]. Available: http://dx.doi.org/10.1109/EUROMICRO.2007.53

[15] P. Stefaniˇ c, D. Kimovski, G. Suciu, and V. Stankovski, "Non-ˇ functional requirements optimisation for multi-tier cloud applications: An early warning system case study," in *2017 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computed, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation*

(SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Aug 2017, pp. 1–8.

[16] D. Kimovski, N. Saurabh, V. Stankovski, and R. Prodan, "Multiobjective middleware for distributed vmi repositories in federated cloud environment," *Scalable Computing: Practice and Experience*, vol. 17, no. 4, pp. 299–312, 2016.

[17] A. W. Services, "Amazon Elastic Container Service Developer Guide," 2014. [Online]. Available: https://docs.aws.amazon.com/AmazonECS/latest/developerguide/ecs-dg.pdf

[18] N. A. J. Durillo, J. J., "jmetal: A java framework for multi-objective optimization," *Advances in Engineering Software*, vol. 42, no. 10, pp. 760–771, 2011.

[19] GitHub, "Argoproj," online; accessed 6 May 2019. [Online]. Available: https://argoproj.github.io/argo/

[20] P. Stefanic, M. Cigale, A. C. Jones, L. Knight, I. Taylor, C. Istrate, G. Suciu, A. Ulisses, V. Stankovski, S. Taherizadeh, G. Flores Salado, S. Koulouzis, P. Martin, and Z. Zhao, "SWITCH workbench: A novel approach for the development and deployment of time-critical microservice-based cloud-native applications," *Accepted to Future Generation Computer Systems*.