TAP: A Task Allocation Platform for the EU FP7 PANACEA Project

Erol Gelenbe and Lan Wang

Intelligent Systems and Networks,
Department of Electrical and Electronic Engineering,
Imperial College, London SW72AZ, UK

Abstract. The EU FP7 PANACEA project has designed a QoS driven smart Task Allocation Platform for varied QoS objectives in the Cloud.

The Cloud [9] supports diverse workloads [1, 2, 6, 7, 10] and simple schemes are needed to allocate jobs with satisfactory QoS and low overhead. The PANACEA project's Task Allocation Platform (TAP) uses on-line observation of the servers in a Cloud system to dynamically allocate tasks. TAP is a Linux kernel module which embeds measurement agents into hosts. We illustrate its usage with a smart algorithm inspired by the Cognitive Packet Network (CPN) [3, 5, 8] which uses reinforcement learning [11], and with a "sensible" policy [4] that probabilistically selects the host whose measured QoS is the best. TAP is a practical system shown in Fig. 1 which exploits several different task allocation algorithms such as the two we mention. It is implemented as a Linux kernel module on PCs with Linux OS.

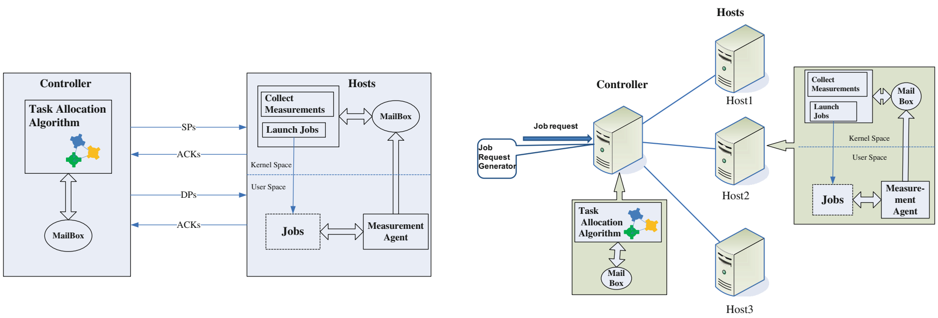
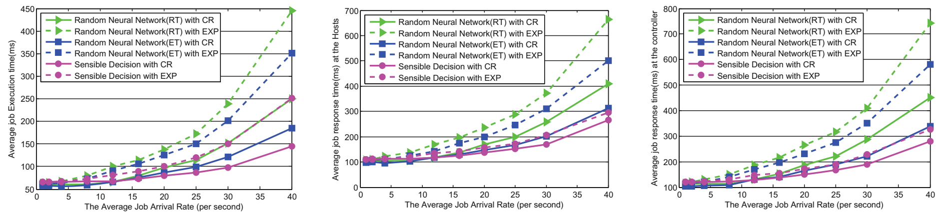


Fig. 1. Architecture of the Task Allocation Platform (left) and its test-bed (right).

A synthetic benchmark is generated, and jobs are sent at fixed intervals denoted by CR, or according to a Poisson process with a fixed rate denoted by EXP. The QoS goals used here are either (i) the minimization of either the execution time on the host, or (ii) the minimization of the response time at TAP, which includes the message sent to activate the job at a host and the time it takes for an ACK to provide information back to TAP. The CPN based

A. Celesti and P. Leitner (Eds.): ESOCC 2015 Workshops, CCIS 567, pp. 425–426, 2016.
DOI: 10.1007/978-3-319-33313-7



(a) (b) (c)

Fig. 2. Average job execution time and average job response time at the hosts, and average job response time at TAP for different job arrival rates.

scheme was tested with both (i) and (ii), whereas the sensible decision approach only used (ii). The experiments are carried out for average job arrival rates of $1, 2, 4, 8, 12, 16, 20, 25, 30, 40$ jobs/sec with each experiment lasting 5 mins. As the average job arrival rates grows, the sensible decision algorithm outperforms the RNN, as shown in Fig. 2. Also the RNN algorithm with online measurement of job execution time performs better than the RNN with the metric of job response time, and the sensible decision is always best under high job arrival rates.

References

1. Berl, A., Gelenbe, E., Di Girolamo, M., Giuliani, G., De Meer, H., Dang, M.Q., Pentikousis, K.: Energy-efficient cloud computing. Comput. J. **53**(7), 1045–1051 (2010)
2. Delimitrou, C., Kozyrakis, C.: Qos-aware scheduling in heterogeneous datacenters with paragon. ACM Trans. Comput. Syst. **31**(4), 12:1–12:34 (2013). http://doi.acm.org/10.1145/2556583
3. Gelenbe, E.: The first decade of g-networks. Europ. J. Operational Res. **126**(2), 231–232 (2000)
4. Gelenbe, E.: Sensible decisions based on qos. CMS **1**(1), 1–14 (2003)
5. Gelenbe, E.: Steps toward self-aware networks. Commun. ACM **52**(7), 66–75 (2009)
6. Gelenbe, E., Lent, R.: Energy-qos trade-offs in mobile service selection. Future Internet **5**(2), 128–139 (2013). http://dx.doi.org/10.3390/fi5020128
7. Gelenbe, E., Lent, R.: Optimising server energy consumption and response time. Theor. Appl. Inform. (4), 257–270 (2013)
8. Gelenbe, E., Timotheou, S.: Random neural networks with synchronized interactions. Neural Comput. **20**(9), 2308–2324 (2008). http://dx.doi.org/10.1162/neco.2008.04-07-509
9. Mell, P., Grance, T.: The nist definition of cloud computing. NIST Spec. Publ. 800–145 (2009)
10. Pradeep, P., Shin, K.G., Zhu, X., Uysal, M., Wang, Z., Singhal, S., Merchant, A., Salem, K.: Adaptive control of virtualized resources in utility computing environments. In: Proceedings of the 2nd ACM SIGOPS/EuroSys European Conference on Computer Systems 2007, pp. 289–302. EuroSys 2007, NY, USA (2007). http://doi.acm.org/10.1145/1272996.1273026
11. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press (1998)