Carretero, J., García-Blás, J., Antoniu, G., Petcu, D. (2018). New directions in mobile, hybrid, and heterogeneous clouds for cyberinfrastructures. *Future Generation Computer Systems*, 87, pp. 615-617.

DOI: 10.1016/j.future.2018.05.073

New directions in mobile, hybrid, and heterogeneous clouds for cyberinfrastructures

Jesus Carretero[a], Javier Garcia-Blas[a], Gabriel Antoniu[b], Dana Petcu[c]

[a] *Universidad Carlos III de Madrid, Spain*
[b] *INRIA, France*
[c] *West University of Timisoara, Romania*

Abstract

With the increasing availability of mobile devices and data generated by end-users, scientific instruments and simulations solving many of our most important scientific and engineering problems require powerful solutions. These solutions should provide the whole chain to process data and services from the mobile users to the cloud infrastructure, which must also integrate heterogeneous clouds to provide availability, scalability, and data privacy. This special issue presents research works showing advances on mobile, hybrid, and heterogeneous clouds for modern cyberinfrastructures.

Keywords: cloud computing, hybrid clouds, mobile services, performance, energy

1. Presentation

With the increasing availability of mobile devices and data generated by end-users, scientific instruments and simulations solving many of our most important scientific and engineering problems require powerful solutions, providing the whole chain to process data and services from the mobile users to the cloud infrastructure. Following [1], *cloud computing* can be defined as a computing environment where computing needs by one party can be outsourced to a third party and when need be arise to use the computing power or resources like database or emails, they can be accessed through internet. As a result, computing and data are moving away from external devices, desktops, and portable PCs into large world-wide data centers.

To cope with the increasing scale and heterogeneity, hybrid High Performance Computing (HPC) infrastructures, combining the benefits of several technologies are needed [2]. These solutions are more and more important due to the increasing synergies between cloud computing and data intensive applications. This convergence requirea cyberinfrastructures that must be powerful in a broad sense (computation, storage, I/O capacity, communications, etc) to satisfy the services and data processing requirements from millions of users,

Email addresses: jesus.carretero@uc3m.es (Jesus Carretero), fjblas@inf.uc3m.es (Javier Garcia-Blas), gabriel.antoniu@inria.fr (Gabriel Antoniu), dana.petcu@e-uvt.ro (Dana Petcu)

April 11, 2018

but at the same time, they have to provide both strong connectivity and adaptivity utilities to cope with near future mobile applications [3], which will suffer problems like delays, disconnections, failures, etc. Mobile cloud computing [4] can address these problems by offloading mobile applications on resource providers external to the mobile device.

However, the spectacular increase of mobile and IoT devices, makes unfeasible to send all data and control needed to centralise every aspect of the applications. It is then necessary to integrate hybrid clouds, which can provide support for the whole productivity chain, to provide availability, scalability, and data privacy. Inside this productivity chain, it is important to provide edge computing, which means providing computing and storage capacity at the Internet edge in close proximity to mobile devices or sensors. Edge computing allows to locate compute and storage resources near the needs [5], delivering highly responsive cloud services, scalability, and privacy-policy enforcement for the IoT and mobile devices. In addition, edge computing allows to reduce the load on the network and intermediate nodes that would cause the transmission of all data to centralised data centers. Other new paradigms propose to extend the Cloud Computing paradigm to the edge of the network. Fog Computing [6, 7] is oriented toward IoT devices with low latency and fast responsiveness [6, 7]. Mobile-edge cloud computing [8] is a new paradigm to provide cloud computing capabilities at the edge of pervasive radio access networks in close proximity to mobile users.

This special issue presents research works showing advances on mobile, hybrid, and heterogeneous clouds for cyberinfrastructures, including new platforms, system software enhancements, algorithm design and optimization, programming paradigms and techniques, data processing support in homogeneous and heterogeneous computing systems, tools and environment for MHCC data and computing systems, runtime support for MHCC and performance simulations, measurement, and evaluations. The special issue includes extended versions of the selected papers of CCGrid 2017 conference whose topics fit in the scope of this special issue, but it has been also open to any author, through an open call.

2. Special Issue Contents

This special issue of Future Generation Computer Systems Journal contains papers selected from a set of invited papers extracted from the papers presented in the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid 2017), held in Madrid, Spain, but it also covers papers coming from an open call. CCGrid is a successful series of conferences that serves as the major international forum for presenting and sharing recent research results and technological developments in the fields of Cluster, Cloud and Grid computing.

The special issue received 11 papers, 9 featuring extended versions of papers selected among top ranked papers of CCGrid 2017 and two from the open call. Six of which were selected for publication after going through the Future Generation Computer Systems Journal peer review process. As we show below, the accepted papers cover important aspects of the special issue topics, going from more traditional Cloud Computing topics to Fog Computing aspects.

The paper "Predicting Cloud Performance for HPC Applications Before Deployment" [9] presents a machine-learning methodology to support the user in the selection of the best cloud configuration to run the target HPC workload before deploying it in the cloud. This methodology enables the user to decide if and what to buy before facing the cost of porting and analyzing the application in the cloud. It couples a cloud-performance-prediction model (CP) on the cloud-provider side with a hardware-independent profile-prediction model (PP) on the user-side. PP captures the application-specific scaling behavior. The user profiles the target application while processing small datasets on small machines she (or he) owns and applies machine learning to generate PP to predict the profiles for larger datasets to be processed in the cloud. CP is generated by the cloud provider to learn the relationships between the hardware-independent profile and cloud performance starting from the observations gathered by executing a set of training applications on a set of training cloud configurations. Since the profile data in use is hardware-independent the user and the provider can generate the prediction models independently possibly on heterogeneous machines. The results obtained by applying the prediction models to Fortran-MPI benchmarks show that the resulting relative error is below 12% for CP and 30% for PP. The optimal Pareto front of cloud configurations finally found when maximizing performance and minimizing execution cost on the prediction models is at most 25% away from the actual optimal solutions.

In "Performance Modelling and Verification of Cloud-based Auto-Scaling Policies" [10], Evangelidis et al. addresses the problem of auto-scaling, a key property of cloud computing, and propose a novel approach, based on performance modelling and formal verification, to produce performance guarantees on particular rule-based auto-scaling policies. They demonstrate the usefulness and efficiency of the techniques proposed through a detailed validation process on two public cloud providers, Amazon EC2 and Microsoft Azure, targeting two cloud computing models, Infrastructure as a Service (IaaS) and Platform as a Service (PaaS), respectively. The experimental results show that the modelling process along with the model itself can be very effective in providing the necessary formal reasoning to cloud application owners with respect to the configuration of their auto-scaling policies, and consequently helping them to specify an auto-scaling policy which could minimize the QoS violations.

Prasad et al. present in "RConf(PD): Automated Resource Configuration of Complex Services in the Cloud" [11] a novel system, named RConf(PD), that offers to service providers automatic deployment of component instances for complex services such that the resource utilization at the providers premises is maximized in the presence of customer constraints. The system consists of two key technologies that service providers can choose from. Both, RConf and RConfPD, build on an analytical model based on robust queuing theory to accurately model arbitrary components,and ultimately finds the optimal combination of component instances. Real-world experiments presented for Rconf show that, compared to greedy approaches, RConf provisions 20% less resources in the first place, and can reduce resource wastage on live resources by up to 50% at the same time. The second solution (RConfPD) trades-off some of the optimality of RConf for a computational expense two orders of magnitude below that of RConf to provision time-sensitive services. Based on a primal-dual algorithm framework RConfPD relaxes the optimality constraints of RConf and

removes dominated combinations to determine an approximation for the optimal solution. The evaluation shows that RConfPD allows for instant provisioning (in many cases < 1 ms), while maintaining 80-99% of the solution quality of RConf.

With the explosion of mobile devices, mobile cloud computing paradigm has been proposed to offload mobile client applications to be run remotely in the cloud infrastructure, so that applications may use fewer resources in the user's mobile devices. The paper "A Heterogeneous Mobile Cloud Computing Model for Hybrid Clouds" [12], by Alonso-Monsalve et al., presents a new mobile cloud computing model, in which platforms of volunteer devices provide part of the resources of the cloud, inspired by both volunteer computing and mobile edge computing paradigms. These platforms may be hierarchical, based on the capabilities of the volunteer devices and the requirements of the services provided by the clouds and the orchestration between the volunteer platform and the public, private or hybrid clouds. This model is proposed as an inexpensive solution to different application scenarios that provide benefits in cost savings, elasticity, scalability, load balancing, and efficiency. The evaluation carried out also shows that the proposed model is a feasible solution for cloud services that have a large number of mobile users.

Internet of Things (IoT) is bringing an increasing number of connected devices that have a direct impact on the growth of data and energy-hungry services. These services are relying on Cloud infrastructures for storage and computing capabilities, transforming their architecture into more a distributed one based on edge facilities provided by Internet Service Providers (ISP). Yet, between the IoT device, communication network and Cloud infrastructure, it is unclear which part is the largest in terms of energy consumption. In the paper "End-to-end Energy Models for Edge Cloud-based IoT Platforms: Application to Data Stream Analysis in IoT" [13] Li et al. provide end-to-end energy models for Edge Cloud-based IoT platforms. These models are applied to a concrete scenario: data stream analysis produced by cameras embedded on vehicles. The validation combines measurements on real test-beds running the targeted application and simulations on well-known simulators for studying the scaling-up with an increasing number of IoT devices. The results presented show that the edge cloud part embedding the computing resources consumes 3 times more than the IoT part comprising the IoT devices and the wireless access point.

Fog computing is a recent approach to cope with both mobile and IoT devices using hybrid clouds. Fog-devices are usually physically close to end-devices and have a high-speed connection with cloud servers. However, in this systems, access control is a major challenge, as they only provide good access control between end-devices and the cloud if an ID-based proxy re-encryption scheme is deployed on them. Each file stored on a cloud sever must be encrypted using a symmetric key, and these keys are encrypted by a public master key, which is stored in a fog-device. If an end-device wants to access a file in the cloud, then the fog-device re-encrypts these encapsulated symmetric keys from the master key to the key of the end- device. However, because of the geographic dispersion of fog-devices, they are apt to be attacked by side channel attacks. In the paper "A Leakage-Resilient ID-based Proxy Re-encryption Scheme for Access Control in Fog Computing" [14], the authors propose a leakage-resilient ID-based proxy re-encryption scheme in an auxiliary input model, showing that it can resist the continuous leakage of secret keys caused by side channel attacks. The

implementation results show that the proposed scheme is feasible in practice.

Acknowledgements

We would like to thank all the authors, reviewers, and editors involved in the elaboration of this special issue, including also the reviewers that were involved in the CCGrid 2017 conference. We are especially grateful to Peter Sloot, editor in chief of the Future Generation Computer Systems Journal, for approving this special issue and for his help along the process of its preparation.

References

[1] Y. Jadeja, K. Modi, Cloud computing - concepts, architecture and challenges, in: 2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET), pp. 877–880.

[2] G. Mateescu, W. Gentzsch, C. J. Ribbens, Hybrid computing - where hpc meets grid and cloud computing, Future Generation Computer Systems 27 (2011) 440 – 453.

[3] Z. Sanaei, S. Abolfazli, A. Gani, R. Buyya, Heterogeneity in mobile cloud computing: Taxonomy and open challenges, IEEE Communications Surveys Tutorials 16 (2014) 369–392.

[4] N. Fernando, S. W. Loke, W. Rahayu, Mobile cloud computing: A survey, Future Generation Computer Systems 29 (2013) 84 – 106. Including Special section: AIRCC-NetCoM 2009 and Special section: Clouds and Service-Oriented Architectures.

[5] M. Satyanarayanan, The emergence of edge computing, Computer 50 (2017) 30–39.

[6] F. Bonomi, R. Milito, J. Zhu, S. Addepalli, Fog computing and its role in the internet of things, in: Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing, MCC '12, ACM, New York, NY, USA, 2012, pp. 13–16.

[7] S. Yi, C. Li, Q. Li, A survey of fog computing: Concepts, applications and issues, in: Proceedings of the 2015 Workshop on Mobile Big Data, Mobidata '15, ACM, New York, NY, USA, 2015, pp. 37–42.

[8] X. Chen, L. Jiao, W. Li, X. Fu, Efficient multi-user computation offloading for mobile-edge cloud computing, IEEE/ACM Transactions on Networking 24 (2016) 2795–2808.

[9] G. Mariani, A. Anghel, R. Jongerius, G. Dittmann, Predicting cloud performance for hpc applications before deployment, Future Generation Computer Systems (2017).

[10] A. Evangelidis, D. Parker, R. Bahsoon, Performance modelling and verification of cloud-based auto-scaling policies, Future Generation Computer Systems (2018).

[11] A. S. Prasad, D. Koll, J. O. Iglesias, J. A. Aroca, V. Hilt, X. Fu, Rconf(pd): Automated resource configuration of complex services in the cloud, Future Generation Computer Systems (2018).

[12] S. Alonso, Temporary (2018).

[13] Y. Li, A.-C. Orgerie, I. Rodero, B. L. Amersho, M. Parashar, J.-M. Menaud, End-to-end energy models for edge cloud-based iot platforms: Application to data stream analysis in iot, Future Generation Computer Systems (2017).

[14] Z. Wang, Leakage resilient id-based proxy re-encryption scheme for access control in fog computing, Future Generation Computer Systems (2017).

Authors' Biographies

Jesus Carretero

Jesus Carretero is a Full Professor of Computer Architecture and Technology at Universidad Carlos III de Madrid (Spain), since 2002. His research activity is centered on high-performance computing systems, large-scale distributed systems and real-time systems. He

is Action Chair of the IC1305 COST Action "Network for Sustainable Ultrascale Computing Systems (NESUS)", and he is also currently involved in three other EU projects. Prof. Carretero is Associated Editor of the journal Computer and Electrical Engineering and International Journal of Distributed Sensor Networks and has been guest editor for special issues of journals as International Journal of Parallel Processing, Cluster Computing, Computers and Electrical Engineering, and New Generation Computing. He has been General chair of HPCC 2011 and MUE 2012, and Program Chair of ISPA 2012, EuroMPI 2013, C4Bio 2014, and ESAA 2014. Prof. Carretero is a senior member of the IEEE Computer Society and member of the ACM.

Javier Garcia Blas

Javier Garcia Blas has been a visiting assistant of the University Carlos III of Madrid since 2005. He has cooperated in several projects from various high performance research institutions including HLRS, DKRZ, and Argonne National Laboratory. He is currently involved in three projects funded by European projects. Additionally, he is currently involved in various projects on topics including parallel I/O, cloud computing, and accelerators for high-performance platforms. He has participated in many conference organization committees, and in the last three years he has been Program Chair of EuroMPI 2013, C4Bio 2014, ESAA 2014, CCGrid-Life 2015, and IASDS 2015. He counts with 35 research publications in international journals and conferences.

Dana Petcu

Dana Petcu (Mrs., PhD) is Professor at Computer Science Department and vice-rector responsible with international relationships at West University of Timisoara, scientific manager of its supercomputing center, and CEO of the research spin-off Institute e-Austria Timisoara. Her interest in distributed and parallel computing is reflected in more than two hundred papers about Cloud, Grid, Cluster or HPC computing. She is and was involved in several projects funded by European Commission and other research funding agencies, as coordinator, scientific coordinator, or local team leader. She is chief editor of the open-access journal Scalable Computing: Practice and Experience.

Gabriel Antoniu

Gabriel Antoniu is a Senior Research Scientist at Inria, Rennes. He leads the KerData research team, focusing on storage and I/O management for Big Data processing on scalable infrastructures (clouds, HPC systems). He currently serves as Vice Executive Director of JLESC Joint Inria- Illinois- ANL-BSC-JSC-RIKEN/AICS Laboratory for Extreme-Scale Computing on behalf of Inria. He received his Ph.D. degree in Computer Science in 2001 from ENS Lyon. He leads several international projects in partnership with Microsoft Research, IBM, Argonne National Lab, the University of Illinois at Urbana Champaign, Huawei. He served as Program Chair for the IEEE Cluster conference in 2014 and 2017 and regularly serves as a PC member of major conferences in the area of HPC, cloud computing and Big Data (SC, HPDC, CCGRID, Cluster, Big Data, etc.). He has acted as advisor for 18 PhD theses and has co-authored over 130 international publications in the aforementioned areas.