Keynote

Towards understanding Uncertainty in Cloud Computing

Andrei Tchernykh

CICESE Research Center, Ensenada, Baja California, México,
chernykh@cicese.mx
http://usuario.cicese.mx/~chernykh/

Clouds differ from previous computing environments in the way that they introduce a continuous uncertainty into the computational process. The uncertainty becomes the main hassle of cloud computing bringing additional challenges to both end-users and resource providers. Both require waiving habitual computing paradigms, adapting current computing models to this evolution, and designing novel resource management strategies to handle uncertainty in an effective way.

In this talk, we discuss the role of uncertainty in the resource/service provisioning, investment, operational cost, programming models, etc. that have not yet been adequately addressed in the scientific literature. We discuss several major sources of uncertainty: dynamic elasticity, dynamic performance changing, virtualization, loosely coupling application to the infrastructure, among many others. A workload in such an environment is not predictable and can be changed dramatically. It is impossible to get exact knowledge about the system. Parameters such as an effective processor speed, number of available processors, and actual bandwidth are changing over the time. Elastic escalation process has a higher repercussion on the QoS, but adds another factor of uncertainty.

In most existing solutions, it is assumed that VMs are predictable and stable in their performance. In actual cloud infrastructures these assumptions do not hold. Providers might not know the quantity of data and computation required by users. For example, every time when a user requires a status of his e-mail or bank account, it could generate different amount of data and take different time for delivering. A pool of virtualized, dynamically scalable computing resources, storages, software, and services add a new dimension to the problem. The manner in which the service provisioning can be done depends not only on the service property and needed resources, but also users that share resources at the same time, in contrast to dedicated resources governed by a queuing system.

We provide a picture view of uncertainty and its classification in different scenarios from HPC, Grid and Cloud Infrastructures to exascale, and discuss opportunities and challenges of its mitigating.

Andrei Tchernykh is holding a full professor position in computer science at CICESE Research Center, Ensenada, Baja California, Mexico. He is chairing the Parallel Computing Laboratory. He obtained his PhD in 1986 from the Institute of Precision Mechanics and Computer Engineering of the Russian Academy of Sciences, Moscow.

Tchernykh leads a number of research projects and grants in different countries funded by CONACYT, NSF, ANII, Ochoa, INRIA, FNR, UC MEXUS, DAAD, LAFMI, UJF, INPG, REDII, FUMEC, etc. He has delivered more than 70 keynote speeches and invited lectures. He has published approximately 200 papers, and served as a TPC member and general co-chair of more than 240 professional peer reviewed conferences. He has graduated 36 Ph.D. and M.S. students, and served as the External Examiner for Ph.D. programs in Germany, Luxembourg. México, and France. Professor Tchernykh is a founding member of the Mexican Supercomputer Society (RedMexSu).

He is awarded Global Scholars Fellow at Tsinghua University (China), German Academic Exchange Service fellowships at University of Göttingen, Dortmund University, Technische Universität Clausthal, and Severo Ochoa fellowship at Barcelona Supercomputing Center (Spain). He was a Visiting Researcher at Centre de recherché INRIA Lille - Nord Europe (France), Université Grenoble Alpes (France), Luxembourg University (Luxembourg), Moscow Institute of Physics and Technology (Russia), Institut National Polytechnique de Grenoble (France), University of California–Irvine (USA), University of Southern California (USA), Université Joseph Fourier (France), UdelaR (Uruguay), etc.

He has served as a guest editor for Special issues, and member of the editorial boards of several journals, including International Journal of Metaheuristics, Supercomputing Frontiers and Innovations, etc.

He is active in grid and cloud research with a focus on resource optimization, both, theoretical and experimental, uncertainty, scheduling, load balancing, multi-objective optimization, heuristics and meta-heuristics, adaptive resource allocation, and energy-aware algorithms and security (http://usuario.cicese.mx/~chernykh/)**.**