Institution: University of Westminster

Unit of Assessment: UoA11; Computer Science and Informatics

Title of case study: Gateway technologies for high-performance computing in business, industry and science

1. Summary of the impact

Gateway technologies have enhanced the ability of end-users to engage with high-performance computing (HPC) programs on massively distributed computing infrastructures (DCIs) such as clusters, grids and clouds. The technologies are focussed on the needs of business, industry, organisations and communities; enabling them to extract added business and social benefit from custom high-value services running on a wide range of high-performance DCIs. Typically, such services are based on computational workflows tailored to specific business needs. DCIs may comprise resources already owned (eg. clusters) combined with resources rented on a pay-as-you-go basis (eg. clouds). Several companies and organisations worldwide are currently using the technologies.

2. Underpinning research

The Centre for Parallel Computing (CPC) has for many years been researching into the principles underpinning gateways: the set of complementary and interoperating technologies for development and execution of high-performance computing (HPC) and high-throughput computing (HTC) programs on massively distributed computing infrastructures (DCIs). The CPC research team comprises: Professor Gabor Terstyanszky, Dr Tamas Kiss (Reader), Professor Peter Kacsuk, and Professor Stephen Winter. The route to impact has been supported via a series of collaborative research projects including projects funded by the EC Framework Programmes (eg. EDGeS, DEGISCO, EDGI, SHIWA, SCI-BUS, ER-flow, IDGF-SA and CloudSME projects in key research grants, below) and EPSRC in the UK (eg. OGSA, ProSim, Cloud Pilot projects). Adoption outside the academic environment has been a key aim.

Legacy code support (Kacsuk, Kiss, Terstyanszky, Winter; 2003-present). GEMLCA (Grid Execution Management for Legacy Code Applications) was developed to create, run and manage legacy applications on cluster-based service grids. GEMLCA allows application developers to deploy legacy applications on grids without the need for re-engineering any source code: it exposes them as grid services. GEMLCA was integrated into the P-GRADE Gateway implementation, thus allowing legacy applications to be included within workflows.

DCI interoperability (Kacsuk, Kiss, Terstyanszky; 2003-present). The 3G Bridge was developed to provide interoperability between clouds (Amazon, Eucalyptus, OpenStack), desktop grids (BOINC, OurGrid, XtremWeb) and service grids (ARC, gLite, Globus, UNICORE). The bridge enables seamless execution of jobs and workflows across all these DCI types. Additionally, an automated virtual appliance service was elaborated that creates and deploys virtual appliances that encapsulate applications. The virtual appliances can be executed on DCIs that support virtualisation.

Workflow interoperability (Kacsuk, Kiss, Terstyanszky; 2008-present). A coarse-grained interoperability (CGI) method was developed to allow workflows developed under different workflow systems to interoperate. Users can create and run workflows which may incorporate workflows developed for different workflow systems. The team deployed a simulation platform and developed its key component: a workflow repository for managing workflows and their metadata that enables users to create, add, edit and delete workflow metadata. It also offers a wide-range of browsing and searching features.

Application support (Kacsuk, Kiss, Winter; 2008-present). A generic application porting methodology and documentation framework, the Application Development Methodology (ADM), was developed to support migration of applications to DCIs. ADM covers the whole lifecycle of application porting, i.e. from requirements analysis to deployment. ADM is widely used to migrate applications to both desktop and service grids for example in the EDGeS, EDGI, SHIWA and the Cloud Pilot projects.

Science gateways (Kacsuk, 2001-present). Research in graphical development environments has

resulted in P-GRADE – a framework to support the development and execution of programs (jobs or workflows) on DCIs. P-GRADE was conceived by Prof. Peter Kacsuk (who is co-affiliated with SZTAKI Institute, Hungarian Academy of Science) with a view to helping scientist users to engage with the emerging DCIs (including clusters, desktop and service grids and clouds), by providing a common interface onto a wide range of DCI types, and hiding implementation detail. The portal supports workflow specification, submission to a range of DCI types, execution orchestration, and monitoring. The workflow engine, the GUI, and the backend DCI interfaces are core components, to which a number of additional technologies have been added by the CPC team.

3. References to the research

Key publications

Legacy code support:

1. (*) T. Delaitre, T. Kiss, A. Goyeneche, G. Terstyanszky, S. Winter and P. Kacsuk: GEMLCA: Running Legacy Code Applications as Grid Services, in *Journal of Grid Computing,* Vol. 3. No. 1-2. June 2005, Springer Science + Business Media B.V., ISSN: 1570-7873 (paper), pp 75-90, 1572-9814 (Online), http://dx.doi.org/10.1007/s10723-005-9002-8.

DCI interoperability:

2. (*) P. Kacsuk, T. Kiss and G. Sipos: Solving the Grid Interoperability Problem by P-GRADE Portal at Workflow Level, in *Future Generation Computing Systems: International Journal of Grid Computing: Theory, Methods and Applications*, Vol. 24, Issue 7, July 2008, pp 744-751, ISSN 0167-739X doi:10.1016/j.future.2008.02.008.
3. G. Kecskemeti, G. Terstyanszky and P. Kacsuk: Virtual Appliance Size Optimisation with Active Fault Injection, in *IEEE Transactions on Parallel and Distributed Systems*, October 2012, Vol. 23, No. 10, pp. 1983-1995, ISSN 1045-92-19, TPDS-2011-04-0262.R1.

Workflow interoperability:

4. V. Korkhov, D. Krefting, T. Kukla, G. Terstyanszky, M. W. Caan, S.D. Olabarriaga: Exploring Workflow Interoperability for Neuroimage Analysis on the SHIWA Platform, in *Journal of Grid Computing*, published online June 2013, pp. 1-18, DOI:10.1007/s10723-013-9262-7.

Application support:

5. T. Kiss, P. Greenwell, H. Heindl, G. Terstyanszky and N. Weingarten: Parameter Sweep Workflows for Modelling Carbohydrate Recognition, in *Journal of Grid Computing*, Vol. 8, No. 4, 2010, pp 587-601, DOI: 10.1007/s10723-010-9166-8.

Science gateways:

6. (*) P. Kacsuk: P-GRADE portal family for Grid infrastructures, in *Concurrency and Computation: Practice and Experience*, John Wiley and Sons Ltd, Vol. 23, Issue: 3, 2011, pp. 235-245, DOI: 10.1002/cpe.1654.

(*) Indicator of best quality.

Key research grants

- T Kiss, G Terstyanszky, S Winter, H Dagdeviren, *Cloud based Simulation platform for Manufacturing and Engineering (CloudSME)*. Sponsor: European Commission FP7 (Project No. 608886). Duration: 30 months (Start date: 01/07/2013). Value: €750K (Whole consortium: €4.5M)
- G Terstyanszky, T Kiss, S Winter, Building an European Research Community through Interoperable Workflows and Data (ER-flow). Sponsor: European Commission FP7 (Project No. 312579). Duration: 24 months (Start date: 01/09/2013). Value: €137K (Whole consortium: €910K)
- T Kiss, G Terstyanszky, S Winter, International Desktop Grid Federation – Support Project (IDGF-SP). Sponsor: European Commission FP7 (Project No. 312297). Duration: 24 months (Start date: 01/10/2013). Value: €133K (Whole consortium: €860K)
- G Terstyanszky, T Kiss, S Winter, Scientific Gateway-Based User Support (SCI-BUS). Sponsor: European Commission FP7 (Project No. 283481), Duration: 36 months (Start date: 01/06/2011). Value: €349K (Whole consortium: €3.9M)
- T Kiss, S Winter, P Greenwell, G Terstyanszky, Venus-C (Pilot): Protein Molecule Simulation on the Grid. Sponsor: European Commission FP7 (Project No. 261565). Duration: 12 months (Start date: 01/07/2011). €30K (Whole consortium: €3.9M)
- S Winter, G Terstyanszky, T Kiss, P Kacsuk Optimal Scheduling of Scientific Application

Workflows for Cloud-augmented Grid Infrastructures. Sponsor: EPSRC/JISC (Programme: Pilot Projects in Cloud Computing. Project No. EP-I034254-1). Duration: 6 months (Start date: 01/01/2011). Value: £55K

- G Terstyanszky, T Kiss, S Winter, European Desktop Grid Initiative (EDGI). Sponsor: European Commission FP7 (Project No. 261556). Duration: 24 months (Start date: 01/06/2010). €273K (Whole consortium: €1.9M)
- G Terstyanszky, T Kiss, S Winter, Sharing Interoperable Workflows for Large-scale Scientific Simulations on Available DCIs (SHIWA). Sponsor: European Commission FP7 (Project No. 261585). Duration: 24 months (Start date: 01/07/2010). Value: €313K (Whole consortium: €1.8M)
- G Terstyanszky, T Kiss, S Winter, Desktop Grids for International Scientific Collaboration (DEGISCO). Sponsor: European Commission FP7 (Grant Ref. 261561). Duration: 24 months (Start date: 01/06/2010). Value: €284K (Whole consortium: €2.45M)
- G Terstyanszky, T Kiss, P Greenwell, Protein Molecule Simulation on the Grid (ProSim). Sponsor: JISC (ENGAGE: Engaging Research with e-Infrastructure Initiative). Duration: 6 months (Start date: 01/09/2008). Value: £50K
- S Winter, et al, European Research Network on Foundations, Software Infrastructures and Applications for Large-Scale Distributed, Grid and P2P Technologies (CoreGRID). (European Commission FP6 Network of Excellence, Project No. 004265), Duration: 24 months (Start date: 2004). Value: €48K (Whole Project: €9.2M)
- P Kacsuk, A Proposal to Evaluate OGSA/GT3 on a UK multi-site Testbed. Sponsor: EPSRC (Grant Reference GR/S77509/01). Duration: 18 months (Start date: 01/11/2003). Value: £16K (Whole consortium: £119K)

4. Details of the impact

Gateway technologies for HPC to which the CPC have contributed through underpinning research, have created impacts by: (i) allowing users to focus on their HPC and HTC requirements independently of computational infrastructure, through user-friendly portals that liberate user productivity; (ii) facilitating efficient and seamless engagement with a wide range of interoperable computational resources, both hardware and software, that can be owned or leased; and (iii) creating and automatically orchestrating the desired computational workflow on the chosen infrastructure.

This has led to: very efficient use of resources; improved support of products and services within industries and organisations based on the application of HPC; affordable resource usage by businesses, communities and individuals; and better universal understanding of the advantages of HPC and HTC and how to realise them. Thus, self-evidently, the research includes work of direct relevance to the needs of commerce, industry, and to the public and voluntary sectors (including citizens); to the invention of generation of ideas and artefacts such as P-GRADE which have led to substantially improved insights by the end-users; and the use of existing knowledge in experimental development to produce substantially improved products and processes.

Most of the impacts described below have been physically realised via the P-GRADE Gateway implementation, which incorporates all the gateway technologies referred to above.

Economic impacts

In the following companies, the gateway technologies have been used *directly* to improve existing products, processes and services:

(2007-present): Correlation Systems Ltd (Israel).
(2008-present): Simsoft Ltd (Turkey).
(2008-present): E-Group Ltd (Hungary).

In the following companies, the new gateway technologies have been used *within the companies' own research programmes* to improve existing products, processes and services:

(2010-present): Cloudbroker Gmbh and Scale Tools Ltd (Switzerland); 4D Soft Ltd (Hungary); Atos Origin Sociedad Anónima Española (Spain); Simul8 Corporation (UK); MaatG (Spain); Sony Europe Ltd (UK); Worldwide Computer Company Ltd (Charity Engine); MoSGrid (Molecular Simulation Grid), with: Bayer Technology Services GmbH, Origenis GmbH, BioSolveIT GmbH, COSMOlogic GmbH &Co. KG, GETLIG & TAR, TURBOMOLE GmbH, Schrodinger GmbH, and Oracle Deutschland (all Germany); IDGF (Stichting International Desktop Grid Federation), with:

Holografika Ltd (Hungary), Unilever (Global).
(2013-present): CloudSigma AG (Switzerland); ASCOMP GmbH (Switzerland); Ingeniera Y Control Electronico SA (Spain); 2MoRO Solutions SARL (France); Podoactiva SL (Spain); Saker Solutions Ltd (UK); Eurobios (France); Charles Robinson (Cutting Tools) Ltd (UK).

Performance has been improved in companies through highly skilled people having taken up specialist roles that draws on their research in new gateway technologies: Dr C Reynolds (OxFORD Asset Management); Dr Y Zetuny (JP Morgan); Dr Ariel Goyeneche (UBS).

Impacts on public policy and services
In delivering a public service the new gateway technologies have been adopted in the following organisations:
(2010-present): Academic Medical Center (AMC) Amsterdam (Netherlands); Istituto Nazionale di Astrofisica (INAF) (Italy); Stichting AlmereGrid (Netherlands); Erasmus Medical Center (Netherlands); Charite Medical Center, Berlin, (Germany); IDGF (Netherlands); Renderfarm.fi (Finland).

The development of services of benefit to the developing world, in particular desktop grid computing, has been informed by the research:
(2010-present): the technologies have been transferred to institutions in China, Taiwan, Ukraine, Russia, Brazil, and Kazakhstan; IDGF worldwide desktop grid community.

Changes to education have been informed by research:

(2006-present): P-GRADE is used as a teaching tool in other HEIs (eg. Cranfield, Edinburgh, Westminster, HEIs in Hungary) and for PhD training courses in the UK and worldwide. Also, Desktop grids are extracting additional value from existing multiple software license procurements in universities (eg. Westminster, Brunel and Portsmouth). Finally, several National Grid Initiatives deliver public support services for education based on P-GRADE, including those in: UK, Ireland, Croatia, Turkey, Spain, Belgium, Malaysia, Kazakhstan, Switzerland, Australia, HP-SEE (High-Performance Computing Infrastructure for South East European Research Communities), BalticGrid,US Open Science Grid, TeraGrid, Swiss BioGrid.

Impacts on society, culture and creativity
The work of the following charitable organisations has been influenced by the research:

(2010-present): Charity Engine (UK).

Public interest and engagement in science and technology has been stimulated by the research:
(2010-present): IDGF (Netherlands); Academica Sinica (Taiwan); Healthgrid (UK); INTECH Science Centre & Planetarium (UK).

5. Sources to corroborate the impact

Companies

- Co-Founder and CTO, Cloudbroker GmbH, Räffelstrasse 25, CH-8045, Zürich, Switzerland;
- Founder, Simul8 Corporation, 29 Cochrane St, Glasgow, G1 1HL, UK Founder;
- Manager of R&D Projects and Innovation, 2MoRO Solutions, Technopole Izarbel, Cote Basque, 64210 Bidart, France ;
- CEO, gnúbila France, 174, Impasse des Pres d'en Bas, 74370, Argonay, France.

Organisations

- Assistant Professor, Academic Medical Center (AMC), Amsterdam, Netherlands.

Project reports

- SCI-BUS (www.sci-bus.eu); SHIWA (www.shiwa-workflow.eu); EDGI (www.edgi-project.eu); DEGISCO (www.degisco.eu); EDGeS (www.edges-grid.eu).