Mid-Term Theme Report

Title of Theme:	Distributed Programming Abstractions
Theme Leader:	Shantenu Jha
Start Date of Theme:	15 March 2007
Report Date:	05 Nov
Elapsed Time:	7 of 18 (proposed) months
Target Audience:	Distributed Application and Systems Community

Objectives:

(Please restate the original objectives of the theme in bullet point form.)

- Analyse the landscape of Distributed Applications and identify main barriers to Distributed Applications development and deployment
- Understand distributed programming techniques – and provide a critical assessment
- Chart the roadmap for SAGA, i.e. how should a stable high level programming abstraction such as SAGA be provided to the user community
- Identify specific approaches that can be used for application domains (eg Systems Biology) and thus be supported; possibly work on providing implementation(s) of these models and/or abstractions for use
- Provide some insight/recommendations of user requirement to the computer science community (e.g. tool developers) as well as infrastructure developers/deployers.

Chronology of Events:

(A list of meetings and workshops organised and attended, and invited visitor dates with links to any reports, papers, Wikis or other outputs arising. Please include a date, title and *brief* description for each event. Also include the number of attendees and append any exit survey reports – this will be done by admin.)

eSI Opening Public Lecture: 03rd May :

DPA Workshop I (31st May – 02 June):

DPA Authors Meeting (03rd September - 07th September):

DPA Workshop II (31st October – 02 November):

Research Outputs:

(A list of conference presentations and published reports or papers given or in preparation not already itemised above.)

Events/Conferences at which this ideas concerned with/emerging from this theme has been directly discussed:

CompSci2007, Royal Society London, June 2007

ISSGC2007: Introduction to SAGA with a significant section on Distributed Programming Abstractions

CyberInfrastucture Summit, Alberta, Canada, October 2007

OGF21: GridNet and SAGA-RG meeting to discuss service discovery and similar functional extensions to the core (first) SAGA specification.

Grid Today Article: http://www.gridtoday.com/grid/1735208.html

Main Section – Progress to Date:

(A narrative description of the results and any interim conclusions achieved to date, how pursued and how ideas are evolving. This should refer to the chronology/outputs above. In particular:

- Has the work completed to date suggested new opportunities and possible enhancements to the theme?
- Are the original objectives still appropriate or should a re-alignment be considered in the light of progress to date?
- Do you see any potential for follow on activity after the theme ends and if so, of what type e.g. new collaborations, research proposals, extension of the theme itself?)

The scope of Distributed Programming is broad. Thus it was imperative that the theme define clearly the application classes and programming techniques in scope. For example, a large fraction of applications that use distributed infrastructure, such as the National Grid Service (NGS) or the TeraGrid are those that do not explicitly marshal distributed resources (referred to as grid-unaware applications); for such applications, the focus should be on easing the deployment and usability of the distributed resources rather than finding novel and simpler ways of programmatically exposing these distributed resources. However, for applications that need to explicitly marshal resources or those that would benefit from enhanced composibility across resources (referred to as grid-aware applications) the focus has to be on the programmatic interface(s) and composition tools as well as programming models.

The focus of the theme has been to define the landscape of distributed applications, which we think has two primary aspects. The first is an analysis of the features that make distributed applications both more interesting and challenging than say parallel applications or vanilla desktop applications. Secondly, what tools and techniques exist that enable the development and deployment of distributed applications. An integral part of defining the landscape of distributed applications is the logical next step: a gap-analysis of the previously mentioned two components ie.., a critical perspective of the current approaches and what can/should be done better/differently.

Interestingly although distributed applications have been considered for long, the characteristics that make them difficult to develop programmatically are mostly related to features typically associated with Grid infrastructure, which simply put is “the lack of complete control” over the many degrees of freedom. Specifically developing applications while having to handle issues such as complexity of control orchestration and execution management over resources in different administration domains, incomplete information about resource state, availability and resource unreliability. As it is not possible to provide 'knobs and dials' to control these aspects at all levels for every distinct resource, the question becomes very soon one of a trade-off between 'control' and 'complexity'. A corollary to the previous remark is that it is unlikely that there will be widely usable programmatic approaches; it is very improbable that there will be a common abstraction/interface for distributed systems unlike the message passing interface for the message-passing paradigm. More likely there will be many different flavours, and often single applications will involve hybrid programming models and technologies.

There arose from the first workshop, a consensus for the need for a detailed survey of distributed applications. This will serve as an atlas to other 'novel' application developers. The same survey also:

1. Outlines the main programming approaches used for these applications
2. Discusses all potential programming approaches available – both in general and specifically taylored to the applications above

At a formal level the theme is well on track to deliver objectives, ie. A survey of applications and the main programming approaches in use. The major challenge that remains is the understanding of what the critical gaps are. Other questions are: How, if at all, these programming approaches change when performance constraints are tighter; how when Web 2.0 is forcing academic and commercial internet communities to realign their application development efforts, what are the lessons? (in a way, this theme was proposed before the Web 2.0 “bubble” but with exactly the same questions: how can the programming distributed applications be made easier – either by using composition, or the use of component (reusable) frameworks.

Potential for follow on activity: Omer do you want to chip in here with a link to the EU call? And others?

Future Activity:

(Please report on events such as workshops, talks, visits and visitors, already at an advanced stage of planning and implementation.)

Attempts at organizing 'special' events at Edinburgh have taught us that it is very difficult to get the set of people we desire at a mutually convenient time, i.e., scheduling is an NP-hard problem, more so when the objects being scheduled are people with severe constraints/commitments! Thus we have decided to take some of our events to where the desired set of people **are** already congregating! Consequently, we have decided to hold three DPA Theme events away from Edinburgh. This has the added advantage of raising the profile of eSI Themes in general; we believe this is also a significant cost saving strategy too. We will still have a third workshop and couple of focussed meetings at the eSI.

DPA Workshop III, in conjunction with MardiGras Conference (Jan 2008): This will be a day long workshop of invited speakers, who have significant insight and experience to contribute to this field but have The Theme Leader, along with Daniel Katz are the primary organizers of the MardiGras conference, so are able to tailor the event and the participation so as to have maximum impact for the theme. Additionally, a motivation for holding a DPA event (workshop) in the USA is to enable the many people/scientists who would like to contribute/be involved with the Theme are unable to travel to Edinburgh, either due to teaching commitments and/or travel/family constraints.
Accepted speakers are Rich Wolski, John Shalf and Erik Schnetter. There are two panels planned. The first panel will be based upon “homework” assigned to three different workflow tool developers who will be requested to show how their engine/tools would implement a specific problem.
The second panel will involve bringing resource providers together, analysing the set of applications that are run on their systems and inquiring about what can be done to support better a broader range of grid-aware applications...

CCGrid 2008, Lyon: Programming Abstractions Panel Discussion: We are planning on organizing a DPA external event at CCGrid because of the potential for attracting a new group of people who could both benefit from and contribute to the Theme ideas e.g., CoreGrid Advanced Programming Model group/division. We are planning a two hour session – with half the time being allocated to invited presentors and the other half going towards discussions.

Europar 2008 (Workshop): We will organize a workshop at this event, where contributions will be mostly submitted (and thus leading to a high quality peer-reviewed journal publication; we will seek a special edition of a respectable journal e.g., Future Generation Computing Systems)

Another event we are targeting is the **UK e-Science All Hands Meeting**, Edinburgh (September 2008). At the time of writing this report, the situation with the organization/PC is unclear. If this were to happen, this would possibly serve as the Closing Event of the theme (possibly followed by the closing public lecture).

Looking beyond the theme:

The key ideas being proposed in this theme relate very closely to "service-oriented" computing emphasis from the European Commission in the Framework 7 programme. More specifically, results from the DPA will have an impact on: (1) Service-Oriented Knowlege Utilities (SOKU) effort, some of which has shaped the Service Oriented Architectures call (part of Objective 1.2) and the EU Infrastructures initiative; (2) NESSI Technology Platform -- which also has working groups closely related to the DPA theme, particularly focusing on development of application services which can be shared across different business applications. Another European dissemination route being explored by members of the DPA theme is through European "Technical Groups" in the Grid computing areas. It is expected that where gaps have been identified in the DPA theme, a proposal to the European Commission in the "Future and Emerging Technologies" initiative will be submitted in May/June next year. FET is an open call, and focuses on longer term research themes of interest to the computer science and application sciences community.

Any Other Observations:

- Omer Rana: Very soon into the theme, it was realized that the theme did not have a “card carrying computer scientist” as a core-member and was dominated by computational scientists. Whereas the presence of computational scientists is good, it does leave a gap. We are grateful that Omer was willing to step in to provide the 'pure computer science theoretical insight that was missing earlier.
- Scheduling : Prof Peter Coveney as a co-PI on this theme. The PI has been in constant communication with him, but due to Coveney's demanding schedule, all events that have been organized have been at inconvenient dates. Rather than delay the onset of the theme, it was thought prudent that we press ahead with the theme; given that certain EU FP7 calls are over, we anticipate greater availability for direct contribution to the theme.
- A request for an extension to September 30 2008. For both administrative reasons, as well as helps coordinate with several other planned events, which we hope to co-locate at.
