Distributed Programming Abstractions

What are the Challenges for Distributed High-Performance Applications?

Proposers:

Theme Leader: Dr. Shantenu Jha, is an Assistant Research Professor in the Deptartment of Computer Science and a Senior Research Scientist at the Center for Computation and Technology, Louisiana State University. He is also a Visiting Fellow at University College London See attached CV for details.

Theme co-leader (Applications): Prof. Peter Coveney holds a Chair in Physical Chemistry and is Director of the Centre for Computational Science (CCS) within the Department of Chemistry at UCL. Peter Coveney is the PI for the EPSRC RealityGrid Project (2001-2009). He is also co-Director of the UCL e-Science Centre of Excellence. He holds an Honorary Professorship in Computer Science, also at UCL. For more details on Peter Coveney, see: http://ccs.chem.ucl.ac.uk/people/peter/index.shtml

Theme co-leader (Distributed Systems): Prof. Gabrielle Allen, Associate Prof. Department of Computer Science, LSU. Gabrielle Allen is Associate Professor in Computer Science at Louisiana State University, and the Assistant Director for Computing Applications at the Center for Computation & Technology. Gabrielle obtained a PhD in computational astrophysics from Cardiff University in 1993, following undergraduate degrees in mathematics and theoretical physics at Nottingham University and Cambridge University. Before moving to LSU in 2003, Gabrielle was the lead of the computer science area of the Max Planck Institute for Gravitational Physics (AEI) in Potsdam, Germany where she researched and developed techniques for high performance and grid computing. At the AEI, Gabrielle was the lead of the Cactus Code project and a PI for the European GridLab project. See http://www.cct.lsu.edu/g̃allen

1. Theme topic and brief description.

Distributed Programming Abstractions, Models and Infrastructure: In spite of much worldwide effort, the number of scientific applications[1] that can effectively utilize distributed infrastructure remains minuscule. Additionally, there remain only a a handful of applications that have been designed *a priori* for distributed systems, e.g., Wolski's GridSAT. The pressing question that the proposed theme will address is: why in spite of the resources that have gone into the global grid effort, do the previous two statements remain true? Are these symptomatic of a silent killer?

The issues facing the applications community today are very different from those faced five years ago. Earlier it was generally agreed that the major barrier to grid-enabling applications was the status of the infrastructure and simple programming abstractions. There has been progress in the infrastructure - though probably not as much as required - but it has also become obvious that there are additional bottlenecks to the wide-spread adoption of grid applications and the effective utilisation of these distributed resources. The theme will identify the main bottlenecks that exist.

As is commonly known, of all the exponentials in the computing industry, the typical time for bandwidth to double is the smallest – hence the computing environment of tomorrow will not be anything like the stand-alone (serial or parallel) environment of yesterday. The infrastructure is changing drastically and the changes are not confined to just the hardware level. The dynamic infrastructure with often novel features, will make utlising the infrastructure for routine science even more complicated. Thus the evolving infrastructure presents both unique opportunities as well as challenges for applications and application scientists. Of alarming concern however, is that the majority of scientists retain the same mental model of programming applications for distributed infrastructure, as they did for isolated computers. Thanks to many years of experience and research, it is relatively easy to determine which algorithms and data-structures should be used such that application codes scale well across a single platform. But how should scientific applications – both compute intensive and data intensive – be programmed such that they are easily able to utilise a distributed infrastructure? How can computer scientists aid the design and implementation of applications that shields the applications from future trends & changes in computing environments?

It took something as effective as MPI to facilitate a change (in the mindset) to enable scientists to go from serial to parallel architectures. SAGA is proving effective in providing the simple abstractions required and in hiding the heterogeneity of the underlying computational infrastructure. But SAGA does not provide a programming model for distributed applications, high-end or otherwise. SAGA is a necessary step for many, but unfortunately it is not sufficient for all applications. Beyond the top most application layer at which SAGA

[1] The term *application* is an overloaded one (one person's application, maybe another's tool e.g., gridFTP), so it is important to state clearly what we mean by applications: We refer to applications, as the typical scientific programs that end-users design, program and execute.

is effective, what are the other levels at which simple yet powerful programming abstractions can be effective in distributed environments? For example, are workflows the correct abstractions that application developers and programmers should be concerned with? What other abstractions should we be thinking about? Does it even make sense to describe an infrastructure as natural without due consideration to the "programming model"? Which grid computing environments can help overcome these barriers? These issues will be at the heart of what the proposed theme hopes to address. In addition to the proposed theme leaders, we believe we have identified a set of leaders who will help nucleate and then develop these threads of activity within the broader theme.

The theme proposes to systematically explore the main challenges in the design, implementation and effective uptake of distributed applications. The theme will facilitate discussions amongst established leaders from the distributed-systems, applications and computer science domains. We will work hard to ensure that as many of the primary participants agree to spend time at the eSI (we have already received assurances from some) to help build up and follow-through on the discussion agenda and contribute to exploring logical extensions via funding opportunities.

The primary deliverable of the theme will be a report/white paper, presenting the main findings of the workshops etc. and discussing possible mechanisms to address the findings. We are very optimistic that the theme will ignite sufficient research amongst participants, during the life of the theme to produce at least a set of conference and journal papers.

In summary, some of the specific questions that the theme hopes to address are:

- What are the "true" main barriers to wide-spread distributed HPC applications?
- Is there a level other than that at which SAGA operates, at which simple yet powerful programming abstractions can be effective in distributed environments?
- How can computer scientists aid the design and implementation of applications that shields the applications from future trends & changes in computing environments?
- There appears to be also be trend towards domain specific cyberinfrastructure. The pressure to organize around specific domains is understandable. But is this correct evolutionary path? Is there a better way to do?

2. Proposed leader(s)

Jha's specialisation spans computer science and computational science. He is currently a Senior Research Scientist at the CCT, LSU and leads the Computational Biology Research Group. Effective Jan 2007, he will also be an Assistant Research Professor in the Department of Computer Science at LSU. He is also a

Visiting Fellow at the Centre for Computational Sciences, UCL. Jha is currently a co-chair of the SAGA research group of the Open Grid Forum. Jha was the technical lead on the US-UK SPICE project; he is also the receipient of several international awards and prizes. See appended CV for details.

The proposed theme leader (TL) has identified two additional co-leads – one an expert from the application domain (Coveney) and the other (Allen) an expert from the distributed system community with very strong ties to the application community. The three will work collectively to develop the theme scope and workshop(s). Both co-leaders are world leaders in their respective domains. See

3. Relation to e-Science.

The issues that this theme will explore, are not only related to e-Science but are at the very core of e-Science (and the e-Science program). To quote the Prof. Atkinson, the UK e-Science envoy, "the issue of trying to understand what would make better abstractions for programming large-scale applications in distributed (grid) contexts is central to the future of e-Science. Its focus on the nub of how the applications may be better supported is central to much interdisciplanry research that depends on advanced computing. It would clearly lead to insights pertinent to CS research plus a foundation for advances in many application disciplines."

Application areas that would benefit from the outcomes of this theme: We hope that the broadest set of applications will benefit: specific examples are: Systems-Biology, Coastal Modeling, Dynamic Data Driven Applications, computational bio-medical applications, numerical relativity and cosmology, atmospheric physics, geophysics, ab initio material and device modellers. We will try (and have already had great success) in getting experts from many application field to be involved with the theme and come to, at least the early workshops that are trying to characterise the problem. Given the our approach, our impact will more likely be focussed but deep as opposed to broad but shallow; please see Section 6. for details.

Technical areas that would be engaged and developed: There are many ares of contemporary computer science that we anticipate will both want to and need to be engaged and developed: distributed systems, middleware, components based programming model specialists in addition to performance characterisation modelling, problem solving environments and computational infrastructure experts. We outline two specific areas - the first from the program development environment and the latter concerning the runtime/execution enivornment, both of which are important application scientist concerns, but that are not discussed elsewhere in this proposal:

- A Problem Solving Environment (PSE) is a "complete, integrated computing environment for composing, compiling, and running applications in a specific area" [2]. In many ways, a PSE is seen as a mechanism to

[2] *E. Gallopoulos, E. N. Houstis, and J. R. Rice, "Computer as Thinker/Doer: Problem-*

integrate different software construction and management tools, and application specific libraries, within a particular problem domain. Focus on implementing PSEs is based on the observation that previously scientists using computational methods wrote and managed all of their own computer programs – however now computational scientists must use libraries and packages from a variety of sources, and those packages might be written in many different computer languages. Engineers and scientists now have a wide choice of computational modules and systems available, enough so that navigating this large design space has become its own challenge. A survey of 28 different PSEs by Fox, Gannon and Thomas (as part of the Grid Computing Environments WG at the Global Grid Forum) indicate that such environments generally provide "some back-end computational resources, and convenient access to their capabilities". Furthermore, workflow features significantly in both of these descriptions.

- Service Level Agreements (SLAs) ensure that particular application behaviours are being adhered to, and that enough resources are available for the application to execute effectively. SLAs may be used to identify the types of resources that would be most suitable for a particular application area prior to execution. SLAs are also useful when running applications over multiple resources – requiring co-allocation mechanisms to be supported. At run-time the properties identified in an SLA can be monitored to determine if the require quality levels of these properties (such as network latency, CPU usage, memory usage, etc) are being met, which can then be used as a basis to request (the type, duration) of resources and to determine the types of resource demands an application should make (of the underlying infrastructure) in the future. The outstanding issue from the application perspective is how can all this be done without burdening the application development and/or developer?

4. Are there other similar projects to the proposed theme? What would be their relationship involvement in this programme?

There is a body of work on distributed systems and experience, coming from the UK applications community, e.g. from RealityGrid, BioSimGrid, IB, SimDat and some European projects (e.g. DEISA,GridLab), that have at various levels and in different context addressed many of these issues. The lack of a collective approach however, in abstracting or addressing these issues, has led to the perception that individual projects are only exposing the issues of immediate concern to them; we hope the proposed theme will provide the overarching framework to in which to analyse the experience of the individual projects.

In addition there are several ongoing - with which the proposed theme should explore various levels of coupling. For example, the EU's network of excellence

Solving Environments for Computational Science", IEEE Computational Science and Engineering, 1(2), 1994

CoreGRID has a institute on programming models, mostly investigating a Grid Component Model (GCM). Involvement would be via Prof. Kielmann of Vrije Universiteit, Amsterdam. Kielmann has confirmed his intention to be an active/core member through the entire duration of the theme. Also of significance is the The EU XtreemOS[3] project is developing a Linux-based grid operating system. As part of this project, a Grid (OS) API will be designed and implemented. Involvement would also be via Prof. Thilo Kielmann (who is the the PI for the API group within XtreemOS). The other project that we will aim to work closely with is the vGRADS project and will initiate discussions with Prof. Ken Kennedy (via Allen and Wolski)

5. Identify a focus that will ensure the effort is most likely to be productive i.e. a specific test application domain/current unsolved research challenge.

As is universally accepted, there are only a small number of applications that effectively exploit the infrastructure and even fewer number of applications that have been designed for distributed systems. Both computer scientists and application developers will need to change their approaches to address this major deficiency. However no approach that is too drastic and disruptive to the application scientist is likely to be acceptable, hence there needs to be a combination of short-term changes introduced transparently by computer scientist along with longer-term but subtle changes in the way applications are explicitly programmed. These questions along with the other issues that the proposed theme will address, will be laid to test by focussing on the Virtual Physiological Human (or derivatives thereof) – a specific problem in the general area of Systems Biology – which is considered by many[4] to possess the ideal application charactertistics to successfully exploit grids ("Is this the cliched Killer App?" for the Grid), although this hasn't to the best of our knowledge been successfully demonstrated.

6. Please list any people who have agreed to actively collaborate.

The full list of application scientists that we will attempt to have as active collaborators will be determined in the first month of theme (we have identified a subset and have received their explicit support). This is because our we will not be organizing around discipline, but programming and application characteristics, i.e., each of the application types we will explore will have some unique requirements. *It is important to state explicitly that this does not imply that our impact will be limited. It just means that after careful investigation we will target only a handful/subset of application that are representative of their class of applications with common characteristics and facing the same challenges.* We

[3] https://www.xtreemos.org/

[4] http://www.ctwatch.org/quarterly/print.php?p=43 http://tinyurl.com/yf4q4u

will use the taxonomy developed in the seminal paper "Classifying and Enabling Grid Applications"[5]. as the zeroth order starting point. We may possibly extend the classification grid applications – but we will most certainly update the most relevant application examples. This will almost surely be our first plan of action; and the exact areas that we will choose, Based upon the outcome, we will proceed to identify the application areas and thereby specific project/scientists' involvement with the theme to seek.

Some specific application areas that appear to be promising candidates: Dynamically Data Driven Applications; Systems Biology (Goryanin, Gavanagh)[6]; Loosely coupled Hybrid/Coupled Models (Sekiguichi); Novel applications utilizing high-end networks (Peter Clarke, Bill St. Arnaud); Semiconductor transport phenomenon (Asenov, Millar); Applications amenable to Geographically Distributed Domain Decomposition (Boghosian)[7] ; Bioinformatics (Richard Sinnott); Applications requiring heterogenous resources (Wolski). The above just represents, for example, just a first pass at identifying the application areas that we will focus on. With more thought and structured analysis, it is likely that this list may change significantly; hence we have not sought explicit support from leaders of all areas identified (although we have explicit support from a majority of those mentioned above and some e.g. Sinnott have agreed to be active members through out the entire theme). What we can be sure of however, is that what ever the final set of applications that the theme leaders decide to focus on, we will involve the most suitable application scientists working in those areas, i.e., those best placed and most willing to work with computer scientists.

There are some core computer science issues that will need addressing and that will cut across any/all application areas that we eventually pursue. Hence we have identified and sought support from these computer science experts, e.g., grid application programming model experts like Thilo Kielmann, grid-computing environments Geoffery Fox and distributed components (Anthony McCough). Other computer scientists that have agreed to work with us are Omer Rana, Ian Taylor (Cardiff) and Dave Berry (NeSC). **Omer Rana, Daniel S. Katz and Thilo Kielmann, along with the theme leaders will be part of the core group and play critical roles in the charting the course and and organization of the theme.**

Our confidence at being able to able to extend the net in the desired direction stems from the fact that three proposed leaders all have exceptionally strong ties with many different application communities. For example, Peter Coveney heads the UK wide Collaborative Computational Projects (CCP) – the umbrella body for the Computational Science community in the UK. Also, the TL through his work in developing requirements on SAGA, has a deep appreciation of the challenges and concerns that plague the applications community, as does Gabrielle Allen through her work in the EU GridLab project.

[5] 10.1002/0470867167.ch23

[6] Coveney is already part of an effort to understand applications and architecture requirements for Systems Biology. See http://www.ucl.ac.uk/complex/sysbio2006.htm

[7] see http://www.realitygrid.org/publications/GD3.pdf for an elegant discussion on the programming challenges and performance of these applications

Additionally we have explicit support from Dr. Steven Newhouse, Director OMII and Prof. Malcolm Atkinson, as well as from leaders of the most important Grid resource providers, viz., Dr. Sergiu Sanielevici (Applications Director, TeraGrid), Dr. Rob Simmonds (CTO, Westgrid) and Prof. Victor Alessandrini (DEISA). We will definitely work with the UK NGS (Although most many people had indicated agreement when approached verbally or when the broad aim of theme was discussed with them, the number of unconfirmed participants at the time of submission stems from the fact that the final draft of the proposal was available only during the week immediately before SuperComputing 2006 – which is typically the most action-packed week of the year. Whereas the TL takes primary responsibility for having the proposal ready so late on in the cycle, it should be mentioned, that this was due many circumstances out of control. This includes Prof. Satoshi Matsuoka and Dr. Stephen Pickles and Dr. Jikku Venkat, CTO United Devices[8] The core group in addition to the theme leaders, will be comprised of a balance of computer scientists and domain specialists.

7. Sketch of who is probably working in the area, and/or might be interested.

Please see information provided in Section 4, as there is significant overlap.

One person who we have not approached yet but think may prove to be extremely useful is William Fellows of the 451 Group. By having him be involved in the theme – at least through one of the workshops, but hopefully on a more sustained level – we hope to get significant input from the commercial/enterprise computing perspective but distilled through the filter his experience. Of all Industrial Analysts/Consultants, Fellows it the most appropriate, because of his extensive research into Enterprise Applications and the challenges facing the adoption of Grids for enterprises. Most significantly he is the principal author of "Grid Computing – Enabling Applications for Grid Deployment", an influential publication of the 451 Group. Unfortunately, from previous discussions (as SAGA co-chair, Jha arranged a meeting with him in July 2006), having him on-board might possibly require resources, but given the impact that a broader perspective would have, we will explore the possibility. (The "cost factor" is the primary reason why we have not initiated contact.)

8. Identify the current key research challenges(s) in the area

To the best of our knowledge and understanding, there is no single concerted coherent effort, to address the issue of the problems facing application development *across many distinct application domains* . Most if not all attempts focus on addressing specific attempts — as determined by the last successful

[8] United Devices now has major operations in the UK with London becoming the hub for its commercial activities in Europe.

grant! The need to successfully abstract the challenges and requirements across different disciplines is the most critical challenge.

9. What are the plausible outcomes from the theme?

At the very minimum a theme report – at the level of a technical white paper – will be available. Although we say at a very minimum, we stress however that this is also the primary objective. It is very likely, that there be sufficient original material that will emanate from the workshops and meetings to result in peer-reviewed publishable papers – either as part of a conference proceeding or as part of a journal.

The output will be probably theoretical. However, Bill St. Arnaud suggests that it is should be possible (with additional resources), "... in terms of your eSI theme I think real world demonstrations of some new grid applications should be an outcome - in addition to papers and workshops". Any such application, will definitely take additional effort, e.g., the design, coding and implementation of the application, but it would be a truly valuable outcome if the seeds for such an effort were sown during the theme or resulted as a consequence of the theme.

In order to avoid confusion, it is important to mention, that the theme will not attempt to propose a programming model down to the last detail; such work is truly and well beyond the scope of such a theme. The theme should result however, in a very impact publication that will examine the various models in existence, analyse which models are suitable and should be developed further for grid as well as determine the ideal characteristics of a grid programming models (for different application requirements) should be.

10. Sketch of the kind of events (focus/scope) proposed, where they would be held and who would participate.

The theme will be organized around a series of small focussed meetings and a few workshops which will be typically larger and address broader issues.

Workshop 1: It is planned that the first event of this theme will be a 2-3 day workshop with a dual focus. The first will be to learn from application scientists, what *they* feel are the significant barriers to utilizing distributed computing? The theme leaders will try to get the participants to develop their answers from an application programming methodology perspective and try to ensure this does not become a middleware-centric workshop. The second focus of this workshop will build around the experience of resource providers (RP), to discuss the main problems that users of their systems face and what should be done to alleviate problems – both in the short term and long term. The RPs will also help develop a quantitative understanding of the distribution of jobs based upon domain, type and other possibly other characteristics that currently dominate the usage of their systems. The RP will be requested to be forward looking and conjure up the ideal distributed system that their users would like to have and once the users had it, determine how they would utilize it. There is merit in

bringing both the RP and application scientists together for the same workshop, viz., two different sides of the same coin can be explored.

This will be an important workshop to kick-off the theme. In a way, a lot of the scope of what can be, and what should be addressed during the theme will be set by this initial workshop. Hence significant thought and effort will be required so as to get this correct.

This workshop will require about a dozen prominent application scientists, with impeccable track records in using distributed systems for real-end science. The challenge will be as much in coming up with scientists who can deliver insights into their areas, as well as about deriving the right mix of scientists with the appropriate spread of specialties in order to deliver the desired impact. A possible approach will be to first identify the target areas and then determine a suitable representative from those areas. We have already made some progress, as the list of people endorsing contain several of prospective leaders. For the second focus of this workshop, our target participants will be: National Grid Service (Pickles), DEISA (Victor Alessandrini), WestGrid (Simmonds), NAREGI, TeraGrid and Open Science Grid (specific targets for the latter three yet to be determined)

It is envisioned, that the first workshop will be held say the week after the OGF20/EGEE user meeting which is being held at Manchester 7-11 May. Although timed to be around OGF20, the workshop will be held at the eSI, Edinburgh. This is to take advantage of what is expected to be a higher than normal (by OGF standards) attendance at OGF20.

The Applications Developers and Users RG (APPS-RG) of the OGF has in the past addressed, a subset of the issues and questions that the theme aims to address; although the Research Group has played an important role in the community, especially in organizing topical workshops, we believe a lack of sustained effort, a long term strategy and coherent approach may have been primarily responsible for the insufficient impact. We believe that our proposed theme will go a long way in alleviating these problems. However, we would like to build upon the experience of several present and previous members of the ARG – particularly Thilo Kielmann (VU Amsterdam), Ed Seidel (LSU/Albert Einstein Institute) and Simon Cox (Southampton). We believe this theme may help revitalize and refocus the efforts of the ARG.

Workshop II A very insightful document "A Grid Programming Primer"[9][10] was written by members of the erstwhile Advanced Programming Models Research Group. In the introduction to the document, they outlined that:
Grid Programming will require capabilities and properties beyond that of simple sequential programming or even parallel and distributed programming... Regardless of the reasons due which Grid applications will be heterogeneous and dynamic, can a programming model or tool give those heterogeneous resources a common look-and-feel to the programmer, hiding their differences while allowing the programmer explicit con-

[9] http://www.cct.lsu.edu/ gallen/Reports/GridProgrammingPrimer.pdf
[10] A modified version was subsequently Chapter 21 in the Berman, Fox & Hey book

trol over each resource type if necessary?.... If the proper abstraction is used, can such transparency be provided by the run-time system?... To achieve petaflop rates on tightly or loosely coupled grid clusters of gigaflop processors, however, applications will have to allow extremely large granularity or produce over $\approx 10^8$ *way parallelism such that high latencies can be tolerated..... Hence, what programming models, abstractions, tools, or methodologies can be used to reduce the burden (or even enable the management of) massive amounts of parallelism and maintaining performance in a dynamic, heterogeneous environment?... In light of these issues, we must clearly identify where current programming technology is lacking, what new capabilities are required and whether they are best implemented at the language level, at the tool level, or in the run-time system?.... endeavors to identify and investigate programming methodologies that support the effective development of algorithms and codes that perform well in grid environments. A programming methodology or model can be present in many different forms, e.g., a language, a library API, or a tool with extensible functionality. Effective development means that any such methodology should facilitate the entire software lifecycle: design, implementation, debugging, operation, maintenance, etc. Hence, successful programming methodologies should facilitate the effective use of all manner of tools, e.g., compilers, debuggers, performance monitors, etc. Perform well is intended to have the broadest possible interpretation. In a grid, perform well can mean not only the effective use (high utilization) of high-performance resources, but also the flexible composition and management of resources....*

The APM document did an excellent job of surveying the status of programming methodologies as of 2002. But in a way, it did not close the loop, i.e., did not consider many real applications or bring them into the analysis fold. The proposed second workshop of the theme, will be both a logical extension of the thoughts initially just sketched out in this document, as well as re-examine many of the ideas afresh (with the advantage of hindsight and experience over the five-year interval) . We will try to involve all the authors of the original document.

Workshop III: Applications Revisited: Novel Applications scenarios and Emerging Infrastructure. This is the place to include a discussion of novel application scenarios like applications using novel network features (LONI, CA*net4/5 etc.), new whole organism predictive simulations (virtual physiological human), DDDAS - which include coastal modeling – and the challenges that these new novel applications scenarios introduce. It is important to discuss novel application scenarios that rely critically on "advanced, intelligent infrastructure", to keep infrastructure development and application deployment synchronised and compatible.

The CCT runs an annual conference (Mardi Gras conference) in early February of every year, and preparations for the Mardi Gras conference in Feb 2008 have begun. Tentatively it will revolve around *Distributed Applications*, and the challenges and opportunities for *novel* distributed applications in the era of high-speed, high-bandwidth and low-latency optical lightpaths. The synergy with the current proposed theme maybe obvious. In order to maximise international impact and exposure for the eSI Theme, we will explore the possibility of

having the 2008 Mardi Gras meeting as a satellite meeting of the this theme.
Topical Meetings: Whereas workshops will involve a larger number of high-level talks, the topical meetings will be less structured, i.e., more research focussed and open ended. In general, there will be two types of topical meetings: The first type will aim to explore a specific topic/issue in depth. The second type of meeting will essentially be a follow-up to the larger workshops. We envision no more than six and no less than three, focussed meetings will be held over the course of the theme, some of them maybe back-to-back. We anticipate the smaller meeting will involve about half-a-dozen people; the eSI at Edinburgh will be the venue for these smaller focussed meetings.

The first workshop will be followed by a focussed meeting, say a month later, which will unpack many of the compressed ideas that will be arise during the meeting, e.g., a post-mortem of the various presentations arising from applications/domains could lead to analysis framework like: "Is this a plea for better distributed programming tools'?", "Does this problem require a different programming abstraction?", "Are these application scientists using the right resources for the job?". The workshop will serve as the forum to develop the exact framework around which to organize the meeting discussion.

In a classic chicken-and-egg situation, some of the specific issues that we propose to explore in these meetings will depend somewhat on the participants that are available. But some specific issues that we propose to explore in depth are:

- Status of the current distributed programming abstractions - with focus on virtualization techniques ("Each virtualization level comes with a trade-off between abstraction and control") .

- "Workflows as means of scientific programming"? For an application, currently there are two broad types of usage for workflows systems. The first is for managing the overall workflow process of the application at the service level by allowing high-level choreography of services and their execution. This typically involves providing a mechanism for representing the workflow (e.g. BPEL, SCUFL, DAG, etc), specifying it (e.g. GUI, text file etc) and for staging/executing the course-grained distributed entities (services) with their dependencies. The second type of workflow usage involves mapping Grid jobs including data dependencies directly across a set of resources. This typically involves a high-level textual representation (e.g. DAX, DAG, VDLt/x) and, creating the distributed workflow and running of many jobs (with dependencies) and dealing with their monitoring, fault recovery and even in some case, some optimisation. Therefore, one type of workflow deals at the service level (Triana, Taverna); the other at the job level (DAGMAN and PAGASUS). Ian Taylor (who is the co-author with Deelman and Gannon) of an influential new on the subject, has agreed to lead this thread of activity.

- Grid aware versus Grid unaware: What is the extent of functionality that can an application can achieve, and yet be "grid-unaware"? When should

an application be grid aware? There are efforts to construct new approaches (e.g., ParalleX: A New Execution Model for Sustainable HPC Design by Sterling et al) for the next generation HPC computers. Should there be an analogous approach for distributed HPC applications?

- "Its the middleware, Stupid. Or not?"

Wrap-Up Meetings We anticipate that there will be at least one meeting (of a couple of days) to derive the main lessons and conclusions from the theme. The set of participants will be based upon the thrust of the main findings of the theme, the direction post-theme activities will explore as well as being determined from those that have been most active and contributed through the theme. We will strive to ensure that there is an optimal balance between application scientist and computer science focus – not just at the wrap-up meeting, but through the theme.

11. What difference will be generated by running the theme for 6 or 12 months?

The theme leaders believes that a six month interval is too short to fully explore the many issues. For example, there is at least a month of collective research required before the exact agenda for Workshop-1 can be determined. A slower pace – albeit not so slow as to lose momentum and stall – will ensure sufficient time for the ideas to develop as well as to determine optimal use of resources.

12. Is the topic of the theme so specific that it can really all be "tied up" in 6 or 12¡ months time, or should there be some follow-on?

As mentioned, one of the first steps of the theme leader will be to focus the scope of the workshop so as to address a couple of specific questions. Thus the core aims of the theme will be met in the 9-12 month time frame. However like any successful research project, the hope is the theme will open up more questions than it answers! This in itself should be provide enough intellectual fodder for a detailed and specific call to EPSRC (and possibly to the EU Commission and NSF) for further funding. It is the intended that the output of this call might serve as input to several upcoming European FP7, UK and/or US call-for-proposals. For example, we intend for the output of the proposed theme will inform the requirements – infrastructural, application programming and environments – for projects like the Virtual Physiological Human [11]

13. Are there opportunities for co-funding from other sources?

The proposed theme leaders are not aware of any explicit calls or immediate co-funding opportunities at the time of submission. However, all attempts to

[11] http://www.biomedtown.org/biomed_town/STEP/Reception/step-definitions/STEPConference2#Programme

explore co-funding opportunities with other agencies – governmental and industrial will be made.

14. Project plan: Milestones and Resources

The year-long theme will be organized around four quarters, starting in March 2007. The TL will arrive soon thereafter at the eSI to kick-start theme activities. The aim is to have at least one additional core member or co-TL, in residence at the eSI, for the majority of the duration that the TL is at the eSI. The outline of milestones and activities on a quarterly basis is as follows:

01 Mar '07 - 31 May '07	Kick-off. Define scope. Identify applications then complete list of collaborators. Identify speakers and topics for first workshop to be held in May 2007. **Workshop-I**.
01 Jun '07 - 31 Aug '07	Workshop-I follow-up activities. Based on analysis and feedback from Workshop-I, organize first set of **focussed meetings** in early June. Prepare for Workshop-II in August. **Workshop-II**
01 Sep '07 - 30 Nov '07	Half way mark **review**. Add flesh to the skeleton of ideas from the previous workshops, via a couple of targetted **research retreats**. Revisit initial scope, check course, address gaps. Start planning papers, reports and post-theme activities.
01 Dec '07 - 01 Mar' 08	**Workshop-III**. Analysis and final project report. Finalise post-theme activities plan.