Implications of Grid e-Science and CyberInfrastructure for the DoD High Performance Computing Modernization Program

Geoffrey Fox
Marlon Pierce
Indiana University
September 7, 2003

1 Introduction

This white paper summarizes the current state of Grid Systems [Foster99A] [GGF-A] [Berman03A] [GapAnalysis] with particular attention to their possible relevance to HPCMP applications. We particularly wish to emphasize in this report a comprehensive view of Grid technologies and research: as we describe in the next sections, so-called "(meta)computing grids" are a subset of a much broader international research and development trend. In Section 2, we describe different types of Grids and contrast their use and implementation with clusters and massively parallel systems. Section 3 is a brief review of DoD opportunities stressing data services, portals and application integration rather than the "utility and metacomputing" areas where Grids were first proposed. The technology itself is summarized in section 4 in terms of different services linking Grid and Web Consortium approaches. Section 5 enumerates several possible DoD applications including those inside and outside the PET and HPCMP mandate. Section 6 focuses on the IMT PET focus area which has a particularly rich set of possibilities to use Grid technologies.

2 What is a Grid?

Here we try to distinguish four related networked systems

1) *Classic Massively Parallel Machine* such as the IBM SP series. These are a networked collection of nodes with a custom high performance network whose aggregate bandwidth scales proportionally to the number of nodes. The latency for small internode messages is a few microseconds. Good performance on many parallel applications requires ratio of communication times to calculation times that is not much larger than 10 to ensure low communication overheads [Dongarra02A]. The small messages are common in many cases and the low latency is needed to get good efficiency in this case.
2) *Typical cluster* which is similar to an MPP but constructed from commodity components with usually competitive node performance and bandwidth but often substantially poorer latency in the 100-1000 microsecond range.

3) *Computing Grid* is a distributed system of networked computers which can be very heterogeneous; in particular parallel machines and clusters can be nodes of a Grid. Another well-known case is the "Desktop Grid" consisting of "all the desktop machines" either in an Enterprise (the Condor model [Condor]) or in the world (the SETI@Home model [SETI]). Grids are heterogeneous in both computing nodes and networking and can have inter-node latencies of 100-1000 milliseconds as is typical of wide area networks. A geographically localized Grid could have inter-node latencies of around a millisecond. Computing Grids grant remote users the privilege to directly access computing resources.
4) *Information Grid* shown in fig. 1 is a network of computers, data repositories (both file and database) and sensors. Information grids are characterized by their use of metadata models and services to describe, organize, and provide controlled access to scientific data and resources [GapAnalysis]. Metadata is simply "data about resources" and may be used to describe a) characteristics of large permanent scientific data sets; b) ephemeral information such as computing loads on HPC systems; c) feeds for streaming data; and d) information about groups, individuals, projects, and so forth. A related problem in this area is "data provenance" or "intellectual property" descriptions [myGrid-D]. This is used to describe who created or owns a particular piece of data, how was it created, what assumptions were made, what is the quality of the data, and so forth. Information Grids will be a focus of this paper and will be more fully described in subsequent sections.

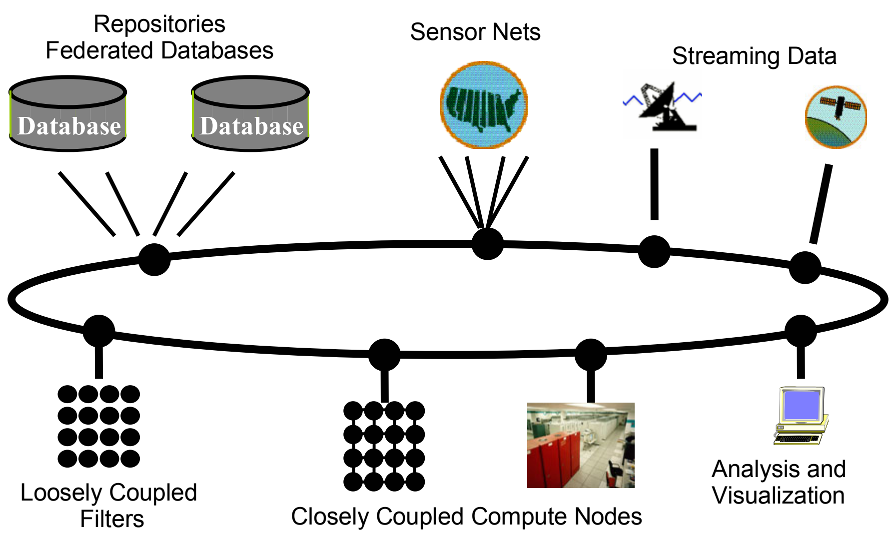

Fig. 1: Information Grid with Sensors, Satellites, databases, high performance computers, clusters and filters (independent machines)

Each of the four networked systems described above has applications for which they are suited and those for which they are less well optimized. The suitability can reflect either functionality and/or cost performance. For example, only class 1) can efficiently execute many parallel applications. On the other hand, the cost per "floating point operation" of class 2) is perhaps half that of class 1) (MPP's) and Grids can usually offer the very best cost performance of all. However this computational performance can only be realized on problems not needing the low latency synchronization and system integration only available in MPP's. This implies that Grids for example cannot easily realize the dream of meta-computing – linking multiple sites together as a single supercomputer – unless the problems can be decomposed into essentially independent tasks. There is not only the problem of very high network latencies (up to 100,000 larger than that of an MPP) but the administratively hard problem of co-scheduling – reserving large blocks of time simultaneously on geographically and administratively distinct machines. There are two important application scenarios where Computing Grids are very appropriate

a) The *unmanaged* or *managed Desktop Grid* where one has a very large set of related jobs which run independently on a pool of desktop class computers. This is familiar from "idle cycle stealing" projects like SETI@Home and in the managed case where the particle physics community expects to keep tens of thousands of computers running continuously each analyzing separate events from CERN's Large Hadron Collider LHC [Condor] [EDG-A] [LCG]. This Grid is arranged hierarchically with central, "national" and "university-level" tiers sharing the load roughly equally. The Grid components are typically each a large cluster. Note this particular application involves substantial data management problems as in full operation around the year 2010 the LHC will produce 10's of petabytes of data per year. This data will demand substantial network bandwidth but typically be staged ahead of time and have a traditional file-based computing structure as opposed to database model of Information Grids.
b) The *seamless access* capability illustrated by the Gateway project from PET [Haupt03A]. Here there is a Grid containing multiple simulation engines and one provides a portal enables submission of a given job at one of the Grid-enabled nodes. As well technology to standardize job entry, staging of files between nodes and client is typically supported in such systems. Further one provides a uniform link to machine and job status information in fashion familiar from the NPACI HotPage resource [HotPage]. Unicore is perhaps the best known seamless access project [Unicore-A].

There is a related application category which bridges computing and information Grids.

c) *Pipelined Grid* resources are an important case as they illustrate the main reason why Information Grids are less sensitive to latency than Computing Grids. Information can typically be streamed from sensors or data repositories so that after a small start-up delay, the large WAN latency is irrelevant. The simplest case of this is a Data/Sensor/Instrument source feeding a computer which itself feeds a user analysis and visualization station. Several early Grid successes fell into this class [Laszewski02A].

Information Grids are well illustrated by the Virtual Observatory and Bioinformatics examples

d) *Virtual Observatories* (VObs) are set up in many fields based around real-time sensors [iVOA]. The initial example comes from astronomy where for example the NVO (National Virtual Observatory) allows access to the results of many different physical observatories linking optical, radio and infrared data. This leads to a new approach to such fields stressing the integration and comparison of results from different data-gathering projects. Earth and environmental science are setting up similar VObs's combining sensor nets, satellites and data repositories as illustrated in fig. 1. A Grid is natural for such applications because not only are the original data-gathering instruments distributed but typically major telescopes each have their own specialized data archive which are also scattered around the world. As a measure of the complexity of an astronomical VObs, this academic field currently has some 10,000 users and 200 repositories worldwide.

e) Bioinformatics has spawned several Grid projects which provide access for the researcher to the growing number of databases in the field. These summarize the results of experiments of many different types and perhaps smaller in volume but much more demanding in the curation requirements [Curation-A] to ensure databases record high quality data. EBI (European Bioinformatics Institute) [EBI] and NCBI (National Center for Biotechnology Information) [NCBI are major organizations providing many of the important databases. As well as the heterogeneous data, this field requires dynamic use of filters (such as the well known BLAST Gene sequence optimizer) which fetch data from the Grid databases and deliver results to the researcher. Virtual observatories also mix Grid computing and data access with image processing as a typical application. In this case each filtering is of the Desktop Grid class as one needs to run multiple instances of the filter on many different data selections.

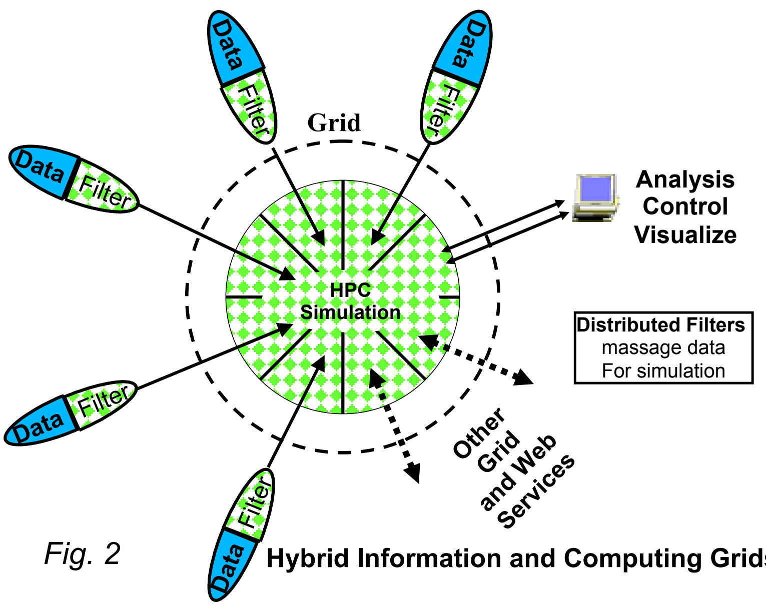

Fig. 2 **Hybrid Information and Computing Grids**

f) Computational chemistry is facing many of the same requirements as bioinformatics. The traditional journal publication approach is far too slow for publishing data to the chemistry community, so distributed data base systems are being developed to house this information. Such data has many metadata requirements, such as who created it and how, where are the associated scientific articles describing the experiment or calculation, and so forth. Such metadata types are termed data provenance, or pedigree. Chemistry information grids also require community curation and annotation. Particular data sets may be "blessed" at some future time, or conversely may be labeled by other researchers as being of dubious quality. The DOE's Collaboratory for Multiscaled Chemical Science (CMCS) is an example of one such project [CMCS].

We have identified four classes of networked computing resources; classic MPP's, clusters, Computing Grids and Information Grids. All are important components in a high performance computing environment and have different application categories where they excel. MPP's and clusters are aimed at tightly coupled problems decomposed with classic parallel computing software and algorithms. These are essential parts of most Grids but the greatest opportunity for Grids is not linking several such supercomputers in real time but rather using one such simulation engine driven by a distributed collection of filters and data resources. This leads to a hybrid Grid architecture shown in fig. 2.

In the following we discuss some possible DoD applications of Grids and follow with a more detailed overview of technology issues underlying the field.

3 Grids and the DOD

The NSF TeraGrid has an architecture similar to that of the HPCMP with around 3-5 major facilities linked to several smaller sites (ETF or Extended Terascale Facility) by a high speed network [TeraGrid]. We see the analogs of MSRCs, distributed centers and the DREN. Although the ambitious Grid implementation of the NSF system is technically impressive, it is not clear how valuable it will be in practice. Correspondingly it is not clear that "Gridifying" the existing HPCMP infrastructure will have major returns in the next few years, since the existing HPCMP Kerberos infrastructure and tools such as the Practical Supercomputing Toolkit [PST] already provide some of the computing grid core functionality for the HPCMP production environment. Rather we think the greatest opportunities lie elsewhere.

First we highlight the use of Grid technology to link applications together; this is code coupling whose importance is clear in CHSSI portfolio projects (SOS, EBE, HIE, SPG, etc.) noted by Bill Zilliox in a recent survey. These will not always need Grid technology but we expect the most robust full featured application integration technologies to come from the Grid workflow field summarized in the following section. CFD/CSM (numerous fluid/structure interactions), CWO/EQM (near shore ocean models), CSM/CCM (multi-scale fracture analysis, nano-materials), CCM/CFD (reactive combustor flow model) are examples of CTA areas coupled in multi-disciplinary simulations that would benefit from workflow. This technology will also support rich data-sets, visualization and computational steering linked to the coupled applications. Portals use Grid technology to build problem solving environments providing users access to these capabilities.

The second major area concerns the information and hybrid grids described above. One could view the OKC as a knowledge (high level information) Grid while FMS, IMT and SIP often require integration of real-time or archived data with simulations. In section 6, we give a commercial aircraft engine (Rolls Royce) example which uses Grids to link sensor nets to real-time system diagnostics to improve maintenance procedures; this appears generalizable to several IMT applications. Generally data management, filtering, interoperability, fusion and provenance are possible intersections of DoD interests with

Grids. Generic DoD areas amenable to Grid technology include Information Security, Systems of Systems, Information Superiority and Decision Dominance, Command and Control and Global Situational Awareness. The *Coalition Agents* Experiment [CoaxGrid] demonstrated real-time command systems built using peer-to-peer Grid infrastructure.

This survey is expanded in sections 5 and 6.

4 Grid Technologies and Capabilities

The Grid field is moving very rapidly and we have provided recent snapshots of its technology, projects and applications in a book and a report summarizing the experiences of the UK e-Science project [UKeS-A] in using existing Grid systems such as that from the Globus team. Several US government agencies have major Grid initiatives including the Information Power Grid from NASA [IPG] and the Science Grid from DoE [Johnston03A]. NSF has recently proposed a CyberInfrastructure initiative which embodies the Grid technical vision described in section 2 combined with the model of science and engineering with collaborative interdisciplinary distributed teams [NSF03A] – the Collaboratory concept introduced by Bill Wulf over a decade ago [Wulf89]. The e-Science project in the United Kingdom has this vision and has made substantial progress with pilot projects and aims at "production deployment" in 2006. In this spirit, they have proposed a new OMII (Open Middleware Infrastructure Institute) worldwide collaboration to coordinate key software architectures and implementations [GapAnalysis].

Such a major world-wide activity can only be sketched here. There is a layered "service architecture" where services are similar to the older distributed object model but with a looser coupling allowing greater robustness and better scaling. The Remote Procedure Call (RPC) at the heart of Java (RMI) and CORBA is replaced by a simpler asynchronous messaging model. Current Grids have replaced the original Globus model by a W3C Web service infrastructure that is optimized for the high performance requirements of Grids. There are some key services we can mention:

a) *Security* is particularly important for heterogeneous distributed systems and essential for e-commerce applications of Grid (Web Service) technologies. Grids require an extension of the traditional transport level security systems (such as SSL). Transport level security insures safe transmission of messages between two set endpoints. Grids on the other hand may need to pass messages through several intermediate hosts and may need to send messages to more than one end point. Grids thus require message-level security in addition to simple, point to point transport level security mechanisms. Current Public Key and Kerberos capabilities for authentication and authorization may be implemented in a message-based Web Service security model whose message-based model has advantages over previous connection-based schemes.
b) *Workflow* captures "programming the Web or Grid" and encompasses a broad range of approaches with names like "Service Orchestration", "Service or Process Coordination", "Service Conversation", "Web or Grid Scripting", "Application

Integration”, or “Software Bus”. There is growing experience and interest in this area with important industry standards such as WSFL and BPEL4WS [WSFL] [BPEL4WS]. Any DoD application needing to link distributed systems will need workflow. Further we believe this approach can be used to provide robust support for the coupling of codes in interdisciplinary applications. We expect many areas to converge on Grid and Web workflow as the common technology for software integration.

c) *The Semantic Grid* is one particularly interesting approach to metadata providing tools and architectures for annotation, search, reasoning about and access to Grid meta-data [SemanticGrid]. This includes a wide range of important capabilities from descriptions of particular services to information about the status of computers and jobs. The growing use of XML and standards based on this format will increase the importance of metadata architectures such as the Semantic Grid.

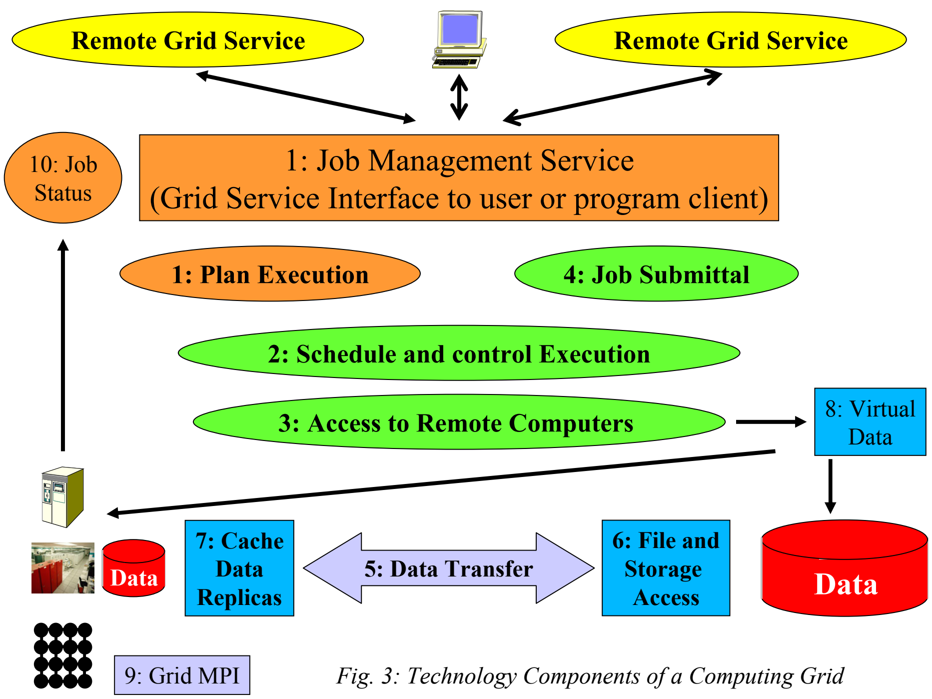

Fig. 3: Technology Components of a Computing Grid

d) *Computing Grids,* shown in fig. 3, are comprised of the many services needed to support distributed computing models. As well as scheduling, planning and job submission familiar from Condor [Condor] and Globus [Globus-A], one needs caching and file management combined of course with the three services described above. File services include Grid flavors of shared file systems (“GridNFS”) and of data transport (GridFTP).

e) *Information Grid Services* include both the OGSA-DAI wrapping of databases as a Web service and the ability to link these to various filters [OGSA-DAI].

f) *Notification Services* provide service control linkage with status and other events propagated between services and between services and portals.

g) *Portals* and their associated services provide the user access to the available Grid services and allow both support of seamless access discussed in section 2 and the construction of problem solving environments [Haupt03A] [Fox03A]. We can expect Grid technologies to be the implementation vehicle of choice for future Problem Solving Environments. The use of handheld devices as the user interface is part of universal access features of current portal work.

h) *Collaboration* or the sharing of Web services unifies areas such as Access Grid and Peer-to-peer networks. Rapid progress is being made and we can again expect greater security and robustness to result for all collaboration tools. Grids, e-Science and CyberInfrastructure are often discussed in terms of virtual organizations (VO). Asynchronous and synchronous collaboration are essential to support VO's.
i) *Network Services* including monitoring, reservation and routing have not received so much attention but should become in future Grids as we need high performance deployment respecting security, resilience and reliability issues.

5 DoD Opportunities

The capabilities discussed in the previous section are receiving major attention from both academic and commercial projects. They appear of importance of many areas of DoD and can importantly enhance several aspects of the HPCMP. We can briefly highlight some general opportunities here with those of special interest to IMT described in the next section:

1) *Seamless Access* to both computers and databases is a near term opportunity building on the PET experience of Indiana, Mississippi State and Texas. This builds off the completed PET ET011 project and involves use of standards for computing and data resources. This is not directly linking resources together but providing users a more uniform access and providing the first step towards more sophisticated Grids; resources with Web and Grid standard interfaces. Seamless access portals are built using the tools shown in fig. 3 for Computing Grids augmented by the growing number of meta-data and Information Grid tools.
2) *Sensor Nets* and their integration with large scale parallel simulation are a natural Grid application that already is in daily use (without perhaps the Grid vernacular) in the CWO focus area. Such integration of real-time data gathering and simulation is applicable to environmental science, solid earth [SERVOGrid] and structural problems where sensor nets can provide real time monitoring and control. There is a general expectation that sensor data will dramatically increase in volume. For instance NASA expects that weather data will grow from 400 megabytes to a petabyte of data gathered each day [ESS02A]. This "data deluge" could lead to new approaches to many fields with a growing importance of data assimilation methods.
3) *Computational Steering* with users controlling simulations with remote portals is an area that has of course already been studied in depth. However Grid services (security, workflow, and notification) provide new approaches and indeed we expect visualization to be reworked to use Grid-based frameworks. The ARL ICE project features both visualization and code coupling and its innovative XML middleware would allow it to be straightforwardly reformulated as a set of Grid services [Clarke02A]. In the Grid approach pre and post processing of HPC jobs would be implemented as the linkage of services using workflow.

4) *Forces Modeling and Simulation (FMS):* Here key technologies HLA (High Level Architecture) and RTI (Run Time Infrastructure) have been developed by DMSO to support FMS simulations. These were very innovative distributed system ideas on their introduction some 5 years ago. However like Java CORBA and COM they probably need to be re-examined in light of the growing importance of Web services. Thus we expect FMS to have growing interest in Grid systems which will be leading the integration of simulation into Web service architectures.
5) *Information Security* is an urgent priority as we find continuing weaknesses in both the Internet and the core operating systems on which it is built. Systematic use of Grid security mechanisms combined with building robust resilient Grid and Web services appears to be part of any approach to information security. This allows us to bypass inevitable flaws in the core infrastructure by only allowing service interactions and engineering these in an Autonomic secure fashion.
6) *Information Superiority* and *Decision Dominance* are at the heart of new military thinking about the conduct of modern warfare. For example, Network-Centric Warfare [Netwarfare] notes that it derives its power from the effective linking or networking of the warfighting enterprise. Joint Vision 2020 [JV2020] emphasises the importance of collecting, processing and disseminating an uninterrupted flow of information while exploiting or denying an adversary's ability to do the same. The UK has recently announced its Network Enabled Capability initiative with the aim of enhancing military capability through the better exploitation of information across the battlespace. As recent world events have shown, multi-national coalitions are playing an increasingly important role in military operations. Indeed, military coalitions are archetypal dynamic virtual organisations that have a limited lifetime, are formed from heterogeneous 'come as you are' elements at short notice, and need secure and partial sharing of information. A common requirement across these programs is the need to inter-operate and integrate heterogeneous distributed systems and to work with large volumes of information and high data rates. We described this earlier under the system of systems concept. In these respects, they could benefit substantially from Grid computing concepts. However, security, resilience, flexibility and cost effectiveness are key considerations for the deployment of military Grids. It is also likely that there will be the need for multiple Grids supporting different aspects of the military enterprise, e.g. 'heavyweight' Grids for imagery data and 'lightweight' ubiquitous Grids running on the PDAs of military commanders in a headquarters—these Grids will need to be interoperable. Currently, there are a number of US military programmes exploring Grid technologies in the context of Network-Centric Warfare, for example Joint Battlespace Infosphere [JBIGrid], Expeditionary Sensor Grid [ExpSensorGrid] and the Fleet Battle Experiments [FleetGrid].
7) The *Coalition Agents* Experiment [CoaxGrid] demonstrated how an agent-based Grid infrastructure could support the construction of a coherent command support system for coalition operations. This illustrates the relevance of peer-to-peer Grids to DoD applications [CoABS-A] [CoABS-B]. More generally we expect future DoD Grids to include the federation of dynamic Wireless Grids (in a war

fighting vehicle for example) to provide more powerful command and control environments.

8) *Global Situational Awareness* is a US defense program whose aim is to "monitor anywhere anytime" with a network of sensors, analysis stations and analysts. This is naturally architected as a Grid but has the constraint that we can't afford to build new weapon systems; rather we must evolve and integrate existing systems. Here one approach is to take each existing system and provide wrappers so that each forms a Grid; then the total DoD environment can be built by federating these existing Grids.

We have given above some general and specific examples of the possible use of Grids in DoD although not all of these are in the PET/HPCMP mandate. We provide below a specific example of the relevance and possible use of Grid technologies to a specific PET/HPCMP functional area, IMT.

6 DoD (HPCMP) IMT Grid Applications

The PET IMT (Integrating Modeling and Testing) component can directly benefit from the Grid technologies described above – especially the information and hybrid Grids presented in figures 1 and 2. Several of capabilities identified as IMT thrusts by PET can be addressed by Grids. We have taken the four thrust areas identified by IMT over the last year and analyzed them from a Grid point of view:

1) *Real Time Modeling:* This includes simulations (using often FMS HLA/RTI technologies) integrated with data and users in the loop. The simulations can be event driven or involve components such as weather generation that require parallel engines. The Virtual Proving Ground [VPG] would be a good candidate for Grid technologies and could use recent commercial initiatives such as that of the openGIS consortium which has issued request for technologies in the area of Web and Grid services for Geographic Information Systems [openGIS].
2) *Data management and Interoperability:* We suggest that an Information Grid is the natural model for IMT as it must encompass distributed dynamic real-time sensors, data repositories recording both previous instrument and the results of engineers analyzing the composite of new and old data. Indiana is preparing a tutorial on high performance Java at the request of ARL who have been using Java middleware to support the staging of data between sensors and repositories. IBM has also identified the critical importance of both robustness and performance in the new generation of server middleware such as Enterprise Javabeans. IBM has announced major Grid and Autonomic (robust, self healing and self-adaptive) computing initiatives addressing such problems [Horn01A].
3) *Integration of Computing and Data:* An interesting commercial example of this is shown in fig. 4 from a collaboration between Rolls-Royce and several UK companies and universities [DAME]. Real-time diagnostic data from aircraft engines (approximately a gigabyte per engine per transatlantic crossing) is fed into data centers. This is filtered through multiple data mining algorithms and compared with previous engine data. Anomalies are flagged and used to enhance the maintenance operations for the airlines using such engines. This uses directly

the architecture of figs. 1 and 2 with data mining being the central high performance algorithm. As noted before, such applications are intrinsically distributed and not sensitive to Grid latencies; data can certainly be pipelined as they flow from aircraft to satellites to ground station repositories and analysis stations.

4) *System of Systems:* This slogan captures the concept that one cannot possibly construct elegant universal systems with clean single standards for every capability. Rather systems will be built leveraging and integrating previous with federation technology such as that pioneered in DoD's HLA approach to modeling and simulation. Grids will support such federation using the new OGSA (Open Grid Service Architecture) to provide interoperability between different existing systems [OGSA]. This federated architecture is important in FMS, Command and Control and IMT. An example of core technology of this type is the new Grid DAI (Data Access and Integration) standard providing a common XML interoperability interface for file and database based repositories and supporting distributed query across multiple federated subsystems [OGSA-DAI].

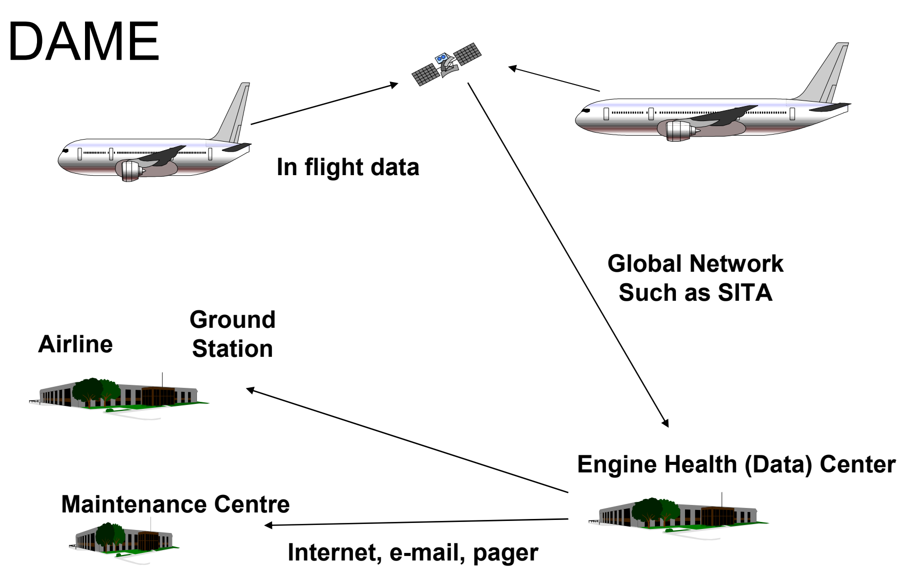

Fig. 4: Distributed Aircraft Maintenance Environment

7 References

1. [Berman03A] *Grid Computing: Making the Global Infrastructure a Reality* edited by Fran Berman, Geoffrey Fox and Tony Hey, John Wiley & Sons, Chichester, England, ISBN 0-470-85319-0, March 2003. http://www.grid2002.org.
2. [BPEL4WS] BPEL4WS: F. Curbera, Y. Goland, J. Klein, F. Leymann, D. Roller, S. Thatte, and S. Weerawarana, *BPEL4WS, Business Process Execution Language for Web Services*, Version 1.0. Available from http://www-106.ibm.com/developerworks/webservices/library/ws-bpel/.
3. [Clarke02A] Jerry A. Clarke and Raju R. Namburu, "A Distributed Computing Environment for Interdisciplinary Applications", Concurrency and Computation: Practice and Experience Vol. 14, Grid Computing environments Special Issue 13-15, pages 1161-1174, 2002.
4. [CMCS] Collaboratory for the Multi-scale Chemical Science (CMCS) http://cmcs.ca.sandia.gov/index.php
5. [CoABS-A] CoABS Grid http://coabs.globalinfotek.com from [CoABS-B]
6. [CoABS-B] Darpa Control of Agent-based Systems CoABS program http://www.darpa.mil/ipto/research/coabs/
7. [CoaxGrid] Coalition Agents Experiment http://www.aiai.ed.ac.uk/project/coax
8. [Condor] Condor Home Page http://www.cs.wisc.edu/condor/condorg/
9. [Curation-A] Seminar sponsored by the Digital Preservation Coalition and the British National Space Centre, Digital Curation: digital archives, libraries, and e-science, London, 19 October 2001, http://www.dpconline.org/graphics/events/digitalarchives.html (Several presentations are available from this link)
10. [DAME] DAME Distributed Aircraft Maintenance Environment project http://www.cs.york.ac.uk/dame/
11. [Dongarra02A] *The Sourcebook of Parallel Computing* edited by Jack Dongarra, Ian Foster, Geoffrey Fox, William Gropp, Ken Kennedy, Linda Torczon, and Andy White, Morgan Kaufmann, November 2002
12. [EBI] EBI: European Bioinformatics Institute http://www.ebi.ac.uk/
13. [EDG-A] European DataGrid EDG http://eu-datagrid.web.cern.ch/eu-datagrid/
14. [ESS02A] Report from the NASA Earth Science Enterprise Computational Technology Requirements Workshop, held April 30 - May 1, 2002 http://esto-doc.gsfc.nasa.gov:8080/documents/Information_Systems/CT/ESE-CT-Workshop/2002/ctreqreport.pdf
15. [ExpSensorGrid] Expeditionary Sensor Grid http://www.nwdc.navy.mil/OperationsHome/CNAN.asp
16. [FleetGrid] Fleet Battle Experiments http://www.nwdc.navy.mil/products/fbe/default.cfm
17. [Foster99A] Foster, I. and Kesselman, C. (eds.). *The Grid: Blueprint for a New Computing Infrastructure.* Morgan Kaufmann, 1999.
18. [Fox03A] Geoffrey Fox, Dennis Gannon and Mary Thomas, *Overview of Grid Computing Environments*, Chapter 20 of [Berman03A]

19. [GapAnalysis] Geoffrey Fox, David Walker, *e-Science Gap Analysis*, June 30 2003. Report UKeS-2003-01, http://www.nesc.ac.uk/technical_papers/UKeS-2003-01/index.html. This report has 390 Grid references and an extensive glossary
20. [GGF-A] Global Grid Forum http://www.gridforum.org
21. [Globus-A] Globus Project http://www.globus.org
22. [Haupt03A] Tom Haupt and Marlon Pierce, *Distributed object-based grid computing environments*, Chapter 30 of [Berman03A]
23. [Horn01A] Paul Horn, IBM, 10/15/2001 presentation at the AGENDA 2001 conference in Scottsdale, AZ, *Autonomic Computing : IBM's Perspective on the State of Information Technology*, http://www.research.ibm.com/autonomic/manifesto/autonomic_computing.pdf
24. [HotPage] NPACI HotPage http://hotpage.npaci.edu/
25. [IPG] NASA Information Power Grid http://www.ipg.nasa.gov/
26. [iVOA] International Virtual Observatory Alliance http://www.ivoa.net/
27. [JBIGrid] Joint Battlespace Infosphere http://www.rl.af.mil/programs/jbi/default.cfm
28. [Johnston03A] Bill Johnston NASA IPG, DoE Science Grid, *Implementing Production Grids*, Chapter 5 of [Berman03A]
29. [JV2020] Joint Vision 2020 http://www.dtic.mil/jointvision
30. [Laszewski02A] Gregor von Laszewski, Mei-Hui Su, Ian Foster, Carl Kesselman, Quasi Real-Time Microtomography Experiments at Photon Sources, in [Dongarra02A]
31. [LCG] LCG: LHC Computing Grid, http://lcg.web.cern.ch/LCG/
32. [myGrid-D] The myGrid Provenance Service http://mygrid.man.ac.uk/myGrid/web/components/ProvenanceData/
33. [NCBI] National Center for Biotechnology Information http://www.ncbi.nlm.nih.gov/
34. [Netwarfare] Network-Centric Warfare http://www.c3i.osd.mil/NCW
35. [NSF03A] Report of the National Science Foundation Blue-Ribbon Advisory Panel, *Revolutionizing Science and Engineering Through Cyberinfrastructure*, http://www.cise.nsf.gov/evnt/reports/toc.htm
36. [OGSA] Open Grid Services Architecture (OGSA) http://www.gridforum.org/ogsi-wg/drafts/ogsa_draft2.9_2002-06-22.pdf
37. [OGSA-DAI] OGSA-DAI Grid and Web database interface http://www.ogsa-dai.org/
38. [openGIS] Open GIS Consortium, Inc. http://www.opengis.org/
39. [PST] The Practical Supercomputing Toolkit. http://www.pstoolkit.org/index.html.
40. [SemanticGrid] Semantic Grid http://www.semanticgrid.org
41. [SERVOGrid] Solid Earth Research Virtual Observatory http://www.servogrid.org
42. [SETI] SETI@Home Internet Computing http://setiathome.ssl.berkeley.edu/
43. [TeraGrid] NSF TeraGrid Project http://www.teragrid.org
44. [UKeS-A] UK e-Science Program http://www.escience-grid.org.uk/

45. [Unicore-A] UNICORE UNiform Interface to COmputing Resources http://www.unicore.de/
46. [VPG] Virtual Proving Ground (VPG) http://vpg.dtc.army.mil
47. [WSFL] WSFL: Web Services Flow Language http://www-3.ibm.com/software/solutions/webservices/pdf/WSFL.pdf
48. [Wulf89] Wulf, William. 1989."The National Collaboratory - A White Paper in Towards a National Collaboratory". Unpublished report of a NSF workshop, Rockefeller University, NY. March 17-18.