Grand Challenges in Computer Science Research in Brazil – 2006 – 2016

Workshop Report – May 8-9, 2006

Authors

- André C. Ponce de Leon F. de Carvalho, USP
- Angelo Brayner, UNIFOR
- Antonio Loureiro, UFMG
- Antonio L. Furtado, PUC-Rio
- Arndt von Staa, PUC-Rio
- Carlos José Pereira de Lucena, PUC-Rio
- Clarisse Sieckenius de Souza, PUC-Rio
- Claudia Maria Bauzer Medeiros, UNICAMP
- Cláudio Leonardo Lucchesi, UNICAMP
- Edmundo Souza e Silva, UFRJ
- Flávio Rech Wagner, UFRGS
- Imre Simon, USP
- Jacques Wainer, UNICAMP
- José Carlos Maldonado, USP
- José Palazzo M. de Oliveira, UFRGS
- Leila Ribeiro, UFRGS
- Luiz Velho, IMPA
- Marcos André Gonçalves, UFMG
- Maria Cecilia Calani Baranauskas, UNICAMP
- Marta Mattoso, UFRJ
- Nívio Ziviani, UFMG
- Philippe Olivier Alexandre Navaux, UFRGS
- Ricardo da Silva Torres, UNICAMP
- Virgílio Augusto Fernandes Almeida, UFMG
- Wagner Meira Jr., UFMG
- Yoshiharu Kohayakawa, USP

Preface

This report summarizes the results of the workshop on "Grand Challenges in Computer Science Research in Brazil - 2006 – 2016", organized in São Paulo on May 8-9, 2006. The Brazilian Computer Society (SBC) promoted the workshop, with support from CAPES and FAPESP. It was attended by 26 Brazilian computer scientists, co-authors of this report, and four invited researchers. The computer scientists were selected by the workshop´s steering committee, based on an analysis of 47 short position papers on grand challenges, submitted from all over Brazil. Selection criteria involved the scientific breadth of each proposal and vision of future perspectives.

The workshop program was divided into three stages. First, each participant gave a short presentation of his/her position paper. Next, there was a debate between two guest speakers – professors Renato Janine Ribeiro and Gilberto Camara – concerning the workshop goals and their views on the challenges to be met. Finally, the attendees were divided into six working groups, which discussed the proposals, consolidating them into five challenges. The workshop was closed by a final session where the results were presented to researchers representing Brazilian financing agencies and R&D companies.

The Brazilian Computer Society has launched the Grand Challenges initiative as the first step of a continuing series of multidisciplinary seminars that will analyze and discuss the workshop's themes in depth. The goals are to foster long-term planning and research in Computer Science in Brazil and to enhance cooperation with other scientific domains.

The Workshop Steering Committee thanks all the researchers who sent their proposals, and in particular those who participated in the workshop. Special thanks are owed to the invited researchers, Carlos Nobre (CPTEC-INPE/ABC – Brazilian Academy of Sciences), Gilberto Câmara (director of INPE), Paulo Artaxo (USP/ABC) and Renato Janine Ribeiro (CAPES, Director of the Evaluation Program) for their contributions and remarks. Finally, the committee thanks the suggestions and the presence, at the closing session, of Carlos Henrique Brito Cruz (Scientific Director, FAPESP), Eratóstenes Ramalho de Araújo (SOFTEX), José Augusto Suruagy Monteiro (Coordinator of the SBC Forum of Brazilian Computer Science Graduate Officers) and Antenor Correa (SEPIN).

May 2006

Carlos José Pereira de Lucena, Claudia Bauzer Medeiros, Cláudio Leonardo Lucchesi, José Carlos Maldonado and Virgílio Augusto Fernandes Almeida

1. Introduction

Computer Science has introduced a revolution in scientific research, being designated as the “third pillar” that supports this research, along with the pillars of theory and experimentation [1]. It thus permeates advances in all knowledge domains. Information Technology, a symbiosis between Computer Science and other areas, mediates new interactions among the sciences, in multiple levels and scales. Indeed, several recent scientific breakthroughs – e.g., in the human genome - result from the work of multidisciplinary teams where cooperation from computer scientists was a fundamental ingredient. Moreover, Computer Science is an essential component to implement and strengthen a country’s scientific, technological, economic and social goals. Thus, long-term planning in the area will have far reaching benefits that go beyond the advancement of Computer Science itself. The workshop on Grand Challenges in Computer Science in Brazil was conceived based on similar initiatives in other countries, which were also motivated by the same observations.

1.1 What is a Grand Challenges Workshop?

A Grand Challenges Workshop is not a traditional scientific event. Rather than aiming at discussion of ongoing research, its goal is to define research issues that will be important in the long run for science and for a country. It must be centered on discussion of scientific issues, and on creative thought, rather than following the usual scenario of scientific conferences, where the focus is on the validation of ideas, proofs and presentation of research results. It is not a conference for “defense of ideas or of personal research agendas” – it should foster collective work towards identifying and characterizing grand research problems. Participants should involve themselves in thinking about problems, rather than on specific disciplines or issues within a topic.

Grand Challenges require multiple views and long-term research endeavors in order to be attacked. They must aim at goals that cannot be reached by short-term planning and incremental projects. The identification of grand research challenges contributes to formulating the so-called *High-Risk-High-Payoff* projects, which have the potential to significantly advance science, with valuable social and technological applications. Some characteristics of Grand Research Challenges include:

1. They must be directed towards significant advances in science, rather than on incremental contributions based on existing results
2. They must present a vision that goes beyond that of a typical grant application for a research project
3. They must be subject to clear and objective evaluation criteria
4. They should be decomposable and amenable to incremental diagnosis, so as to allow changes within the process to achieve them
5. They must be ambitious and visionary, but at the same time feasible within a predefined time frame – in the Brazilian case, 10 years
6. They must be attractive and challenging for scientists, and motivate society as a

whole

7. They originate problems that are multidisciplinary both in nature and solution
8. The topics that motivate them emerge from a consensus of the scientific community, and serve as a long-term scenario for all researchers, regardless of financing policies or issues of the moment

1.2 Similar proposals

Initiatives towards definition of grand research challenges have been undertaken for several scientific domains in countries with long scientific tradition, such as the USA and Great Britain. Examples include

- *Grand Challenges in Environmental Sciences:* organized in the USA by the National Research Council (NRC), by request of the National Science Foundation (NSF), for environmental sciences. One of the guiding principles was to search for inter- and multidisciplinary directions, given that many research problems in this field transcend the traditional barriers among scientific disciplines.

- *Sustainability in the Chemical Industry*: the Grand Challenges and their research needs involved: green chemistry, solvent replacement, improvement in catalysts, renewable energy sources and digital literacy at all levels.

- *The Grand Challenges in Global Health:* initiative conducted on a global scale to try to identify 14 grand challenges that would lead to important advances in the prevention and treatment of diseases that affect the poorest 2 billion inhabitants of the planet.

- *Gordon Research Conferences (GRCs):* international forum for presentation and discussion of research barriers in biological, chemical and physical sciences, and associated technologies.

More specifically, two initiatives have been undertaken in Computer Science to periodically define Grand Research Challenges: the "*Grand Research Challenges in Computing*", in the USA and in the UK. The former, promoted by the National Science Foundation in 2002, produced the following challenges:

1. *Systems you can count on*
2. *A teacher for every learner*
3. *911.net (ubiquitous information systems)*
4. *Augmented cognition*
5. *Conquering complexity*

The UK Computing Research Committee and the British Computer Society produced in 2005 the following challenge list:

1. *In Vivo – In Silico*
2. *Ubiquitous Computing: experience, design and science*
3. *Memories of Life*

4.The Architecture of Brain and Mind
5.Dependable Systems Evolution
6.Journeys in Nonclassical Computation

1.3 The Brazilian Computer Society and the Grand Research Challenges in Brazil: 2006-2016

The Brazilian Computer Society, starting from the model of the international events previously mentioned, organized the Brazilian workshop. The goal was to produce five Grand Challenges for Computer Science in Brazil, together with a specification of a clear and concise view of how to treat the corresponding problem. The formulation of each challenge involved discussion of the following issues: i) specification of the benefits of searching for a solution to the problem, ii) description of how to measure the success of the research undertaken to solve this problem, iii) discussion of difficulties and barriers to achieving success in this research and iv) proposal of actions to be undertaken to meet the challenge in a 10-year period

The five challenges, described in the subsequent sections, were:

1. Management of information over massive volumes of distributed multimedia data
2. Computational modeling of complex systems: artificial, natural, socio-cultural, and human-nature interactions
3. Impacts on Computer Science of the transition from silicon to new technologies
4. Participative and universal access to knowledge for the Brazilian citizen
5. Technological development of quality: dependable, scalable and ubiquitous systems

2. Management of information over massive volumes of distributed multimedia data

Almost everything we see, read, hear, write and measure is collected and made available via computational information systems. To ensure effectiveness and efficiency in processing these data, it is fundamental to develop scalable computational solutions to meet the needs of the applications that will use such data. The goal of this challenge is, thus, to develop solutions for the treatment, retrieval and dissemination of relevant information, both descriptive and narrative, while coping with the exponential growth of multimedia data.

Several factors contribute to the explosive expansion of data sources. First, there is the Internet, through which individuals, enterprises, government agencies and all kinds of institutions are potential content producers, transforming the world into an immense and heterogeneous database, updated in real time, every second, by thousands of individuals. Another critical factor that contributes to this exponential data generation is the appearance of devices that capture new kinds of extremely complex data – e.g., from satellites, microsensors or telescopes, to video cameras that capture human interactions or devices that record brain activity given specific stimuli. Additionally, scientists and researchers generate secondary data during their experiments and access to computational services in a wide range of fields such as earth sciences, astronomy, bioinformatics, medicine or social sciences. These data are of distinct types (for instance, sound, video, text and analogic data to be subsequently digitized) and accessed using different capture and measurement units (e.g., bits, interactions among people, specimen collections in nature), in widely diverse temporal and spatial scales.

This immense heterogeneous data set, distributed all over the planet, needs to be processed, stored and made available to enable information extraction, for a wide range of users, in scalable solutions. This demands, among others, research in new techniques and methods to manage, integrate, index and recover data and information, extracting content from sound or images. Furthermore, it presents challenges in data security and preservation. Moreover, given the constant evolution in information technology, it is necessary to ensure that the stored data will continue to be available and "understandable" throughout the years – that is, ensuring durability and long-term access. Still other research factors to consider involve the modeling of large multimedia data sets, distinct input and output multimodal mechanisms, algorithms and structures for optimized data and information processing, querying and summarization, and an increasing level of support to distinct application and user profiles and requirements.

To manage these large volumes of distributed data and information there is, in addition, the need to efficiently exploit all kinds of parallelism: from the chip level, nowadays with multicore processors, through the level of nodes of high performance networks and clusters, to the grid level, communicating heterogeneous clusters via classical Internet-like networks.

The great challenge is the integration of all these fields, towards applications that can contribute to the country's socio-economic and cultural contexts. Although each of these areas already benefits from (sometimes incipient) research results, there is a lack of proposals that consider them from an integrated point of view.

This challenge is important because, besides stimulating research in basic areas in Computer Science, it fosters their integration, which in turn can contribute to the development of countless key applications for several sectors of society. Examples are: content production for educational activities (*e-learning*), efficient management of information to support electronic government activities (*e-gov*), mining of inter-related data sets to support scientific research (*e-science*), generation of relevant information for medical diagnosis at a distance (*telemedicine*), digital libraries and digital entertainment.

The latter is a good example of the issues involved, and has been progressively gaining pedagogical, economic and social relevance. It is based, among others, on so-called "narrative information" - a modality of information that is little considered in Computer Science, as opposed to descriptive information, typically available in databases. This domain covers practical issues in both fiction and management information systems. For instance, in the latter, it allows simulation, decision-making and training of managers; in the former, it supports story creation and storytelling, electronic games and content production for interactive digital TV. Solutions to these issues involve work in conceptual modeling of genres, development of methods and systems of plot manipulation and design, and experimenting with the requirements of the different application domains.

Briefly, some of the big technical and scientific problems that must be considered to face this challenge include:

- Reduction (abstraction and summarization) of massive data volumes, by means of computational modeling, simulation and others;

- Definition and use of the notion of context for information retrieval, considering factors such as user location, profile and requirements, among others;

- Design and implementation of multimodal content descriptors and of algorithms to extract and index these descriptors, to support multimodal mining;

- Use of distributed and dynamic indexing *peer-to-peer* structures;

- Studies in models and mechanisms to integrate data characterized by a large degree of heterogeneity;

- Treatment, while storing and retrieving data, of factors inherent to the heterogeneity in their capture, such as cultural and temporal issues, but also technological characteristics of captors, such as sensors, cell phones, PDAs, among others;

- Study of alternative means of presenting information, including new kinds of interfaces;

- Consideration of data availability and validity, and intellectual property;
- Formulation of conceptual models to specify genres or domains involved in digital entertainment applications, development of methods and implementation of systems to manipulate plots and their experimental use in various applications;
- Study of adaptable and intelligent infrastructures to process distributed information;
- Development of models and techniques to ensure data and information persistence over long time periods, for historical archiving;
- Development of models, structures, interfaces and algorithms for construction of large distributed digital libraries, to manage multimedia information.

3. Computational modeling of complex systems: artificial, natural, socio-cultural, and human-nature interactions

The term *Computational Science,* coined to contrast with *Computer Science,* is being used all over the world to designate computational models, algorithms and tools to solve complex systems of distinct natures. This kind of research makes it possible, for instance, to study the performance of large computational systems without the need to implement them (e.g., by simulating the behavior of computer networks with millions of nodes). Furthermore, it allows scientists in all domains to investigate problems that until recently were not amenable to analysis – whether due to the volume of data manipulated, or to the absence of analytical solutions, or even to the unfeasibility of studying them in laboratory environments. Examples of such problems are studies in genomics, biochemical processes, particle physics or social interactions with millions of participants (e.g., on the Web or in virtual digital communities). Computational modeling and simulation also allow cost reductions and advances in the industrial sector, through replacing the need for constructing costly buildings and infrastructure by the execution of virtual experiments. The goal of this challenge is to create, evaluate, modify, compose, manage and exploit computational models for a variety of domains and applications, ranging from experiments on artificial systems through to those concerning natural or social interactions.

Mathematical modeling of phenomena is based on well-established principles (for instance, in physics, chemistry, statistics or mathematics); its complexity lies in solving intricate and large systems of equations, which are most of the time already known. On the other hand, in computational modeling the model itself is one of the products of the research. It involves, among other requirements, a large set of algorithms and simulation techniques, data manipulation and mining. The model, in this context, is a computational process that filters, transforms, merges and generates data. This often requires cooperation among scientists in Computer Science and in other domains. Normally, computational modeling research includes uncertainty about the models themselves, since they involve a large number of parameters to be exploited and adjusted.

The complexity of this kind of research grows with the increase in data volume and/or variables to be considered. Another complicating factor is the frequent need for combining several knowledge domains. A typical example thereof, adapted from [1], is the following, "The study of the environmental and biological bases of respiratory diseases requires a complex interdisciplinary modeling effort that couples social science and public health data and experiences with fluid dynamics models of airflow and inhalants (smoke, allergens). It requires, furthermore, materials models of surface properties, biophysics models of cilia and their movements for ejecting foreign materials, and biological models of the genetic susceptibility to disease." Yet another example, mentioned by Câmara [2], shows the need to integrate distinct models – climate change, human and natural systems, socio-economic development and gases and pollutant emissions - while studying environmental problems and global warming.

This field has attracted increasing attention, due to the economic and social benefits associated. Several factors contribute to advances in this kind of research. One is the growing availability of data collected on natural or artificial phenomena, thanks to the dissemination of sensors, whose networks pose by themselves several research challenges. Another factor is the evolution in hardware and software systems, which now allow complex computations, thanks to the adoption of parallel processing and grids – sometimes classified as *large scale Computer Science*. Scientific advances in other knowledge domains, from engineering and natural sciences to the social and human sciences, are also contributing to new and refined models and algorithms.

If all of these factors propel research in computational modeling of complex systems, they also bring up problems for this research due to the massive data volumes generated, and to the growing processing complexity demanded. For instance, satellites and other kinds of sensors already daily generate petabytes (10**15) of data, demanding research in data summarization techniques. Web search engines, such as Google, require a few hundred thousand computers working simultaneously to be able to answer all queries within an acceptable time frame, therefore prompting research in network performance simulation. Studies on interactions of the cells of the human body foresee a possible 10**18 distinct illnesses. All these examples also demand first class research on advanced high performance computer architectures. Thus, at the same time that *Computational Science* has opened up the possibility of research in several heretofore unimagined domains, the advances brought up by work in such new areas generate research challenges in Computer Science.

Computational modeling involves several levels of specific challenges, such as:

- Real time processing of very high speed data streams, generated by thousands of sensors – e.g., in the study of natural catastrophes (such as floods) or in urban emergency evacuation systems given artificial disasters (such as nuclear events). This demands research in, for instance, distributed processing, new database architectures, or networks to support such data streams;

- Design of new requirement elicitation techniques and novel kinds of algorithms and mechanisms to collect and process data, so as to capture variables on social and socio-cultural interactions;

- Devising storage structures to record the computational models and the factors associated with their tuning and execution in a parallel and distributed environment;

- Development of tools for the collaborative construction of models, supporting their execution and modification in real time, so that the model can react to changes in the world while it is executed;

- Creation of new algorithms and techniques in scientific visualization, to support visually capturing the complexity of modeled objects and their interactions – for instance, to help understand the dynamics of a tornado or the evolution of erosion caused by improper human occupation of a region;

- Research on the impact, in software engineering, of the collaboration between computer scientists and those scientists in other domains;

- Management of problems that arise from the increase in scalability and dimensionality (number of variables in a problem), which contributes to exponential processing times;

- Adoption of parallel processing involving heterogeneous resources, such as in computational grids;

- Studies in extensible multimodal interfaces to help in understanding the phenomena being modeled, and to facilitate the dynamic configuration of the models; and

- Integration of algorithms, data structures and models that have been created by distinct scientific domains using their own methodologies.

4. Impacts on Computer Science of the transition from silicon to new technologies.

There is a growing perception that Moore's law[1] will soon no longer hold, given that technology is getting closer to the atom's physical limits. The increase in processing speed is being obtained at the cost of miniaturization, packing ever larger numbers of components into a chip. This, in turn, is augmenting the heat generated and the danger of having components interfere with each other. Several research efforts are thus concerned with developing new technologies to replace or complement silicon. The new proposals – for instance, quantum or biological computing – will require radical changes in the way in which we conceive Computer Science. The goal of this challenge is to analyze the transformations in Computer Science research and development as a consequence of the transition to new kinds of processing paradigms.

All technologies that aim to compensate the limitations of silicon suffer from the same drawbacks – though they have large integration capacity, and allow considerable parallelism, they are slower than the present technologies. Moreover, all present low reliability and high sensibility to noise and to faults, problems that so far have not arisen in (silicon) digital design. Such problems, however, will also materialize in the silicon world, due to increasing transistor miniaturization.

The microprocessor industry has come to terms with the idea that it will no longer be possible to improve performance by augmenting the operation frequency of processors. In fact, the last two years have witnessed the end of the superscalar processor architecture as the sole solution for increasing processing speed, due mainly to its inability to dissipate enough energy at full speed. Multiprocessor solutions have thus appeared as an answer to performance needs. The first approach to this problem was to use the *Hyper-Threading* technique and, from 2005 onwards, the so-called *Dual Core* technology. The tendency adopted by industry is towards the so-called *multicore* architectures, where a large number of highly optimized scalar processors, running at a lower frequency, are connected to a memory hierarchy based on transactional memories, in which applications are developed using a thread-level programming model. Major research challenges in this context include designing new architectures that can provide support to lock-free, transaction-based operations and new operating system and compiler techniques to support such systems.

In the near future, chips will have hundreds of processors, which vary in type (RISC, DSP, VLIW, SIMD, etc.), given the need for power-efficient solutions. Furthermore, silicon-based processors will interact with non-conventional processors, based in quantum or biological computing. Processor heterogeneity and interactions are other examples of issues to be examined.

What are the consequences of this hardware evolution? When speaking of *Computer Science*, we generally associate the notion of program to that of the concrete

[1]Gordon Moore stated that processing speed would double every 18 months, while costs would remain constant

machine in which such programs execute – the silicon chip computers. However, the theoretical foundations of Computer Science do not depend on the physical machines. Computers are just a possible implementation of a machine that is able to *compute*. There exist several proposals of other devices that can execute this work; up to now, such proposals have remained in the realm of academia. The advances of molecular biology can change this panorama, providing Computer Science with a completely different machine: a *biological machine.* Living organisms are reactive systems: given appropriate stimuli (environmental conditions and the presence or absence of certain substances), cells react, producing the proteins/enzymes to respond to the stimuli. The DNA in the nuclei of our cells contains precise information about sequences of bases that generate each protein, indicating when and how these proteins should be produced. In this analogy, the DNA can be seen as the software (encapsulating data and programs), while the cells are the biological hardware. A theoretical model of this biological machine can serve as the basis for the emerging area of Bionanotechnology, providing techniques to design and construct *biological programs.* However, biological computing is suited to massively parallel systems and prone to faults. For instance, in biology, a system is composed of thousands of "processors" (the cells) executing the same task, and the behavior of the system is ensured by the fact that most of these processors will act "appropriately", thus providing a backup for faulty cells. Biological computing thus requires enhanced fault tolerance mechanisms.

In this scenario, and given the demand for new products, there must be a change in the methodologies for hardware and software design. In the present model, software designers use a high level abstraction of the hardware in which the software will execute, without concerning themselves with low-level physical constraints. This will no longer apply: the designer will have to take advantage of massive parallelism, while moreover considering power consumption and reliability. This would seem to indicate the need to discard the practice of designing software based on hardware abstractions, in favor of low-level considerations.

On the other hand, one cannot ignore the growing computational complexity of the problems faced by our society. This seems to indicate the need to work in an apparently opposite direction, that of designing software using very high abstraction levels and increasing amounts of automation, in a scenario of high competitiveness where the *time-to-market* is very short. This brings up the problem of conciliating two opposing issues. On one side are the automation and abstraction required to deal with the complexity and the *time-to-market* requirements of the design of large multiprocessed software. On the other side lies the need for optimization in software and hardware co-design, to attain power efficiency and to compensate for the shortcomings of the low reliability of the new technologies.

Obviously, this new scenario demands a complete adaptation of models and design techniques in several traditional Computer Science fields. In special, new paradigms must be found in several research areas. In Software Engineering, for instance, there will be the need to find the appropriate abstractions to capture low level physical optimization–or to develop new testing and fault tolerance techniques. Operating systems will need to encompass tasks such as controlling power consumption and alternate between hardware and software solutions to execute tasks, in order to minimize energy dissipation. New compilers will require more visibility of non-standard hardware devices. Research in parallel processing will demand combining work in the field of energy management with studies on high performance, and develop new languages to fully exploit parallelism in this context. In this scenario, every computer system will be seen as an embarked system, which can potentially be interconnected with all other resources in a pervasive environment.

At the same time, one must consider the increasing complexity of parallel platforms and highly distributed systems. Future parallel systems will have hundreds of thousands of processors working simultaneously. Thus, another challenge is the development of software that can be scalable to this level, taking full advantage of the architecture to support the appropriate concurrency requirements. Last but not least, most hardware for embarked and pervasive applications will be battery-powered. These batteries still have limited capacity, thereby provoking power bottlenecks.

Summing up, some of the big technological and scientific problems to be faced in this challenge are:

- Conduct research in Software Engineering to support requirements at all abstraction levels. These abstractions should capture the verticality needed by physical optimization of computational systems, ranging from low-level issues such as performance and power consumption, to the high-level abstractions needed to support the growing automation of software development processes.

- Create a more general notion of component-based development, so as to also take non-functional characteristics into account, such as performance, power consumption and dependability.

- Develop new testing and verification techniques, to account for permanent and transitory failures, and for the interactions among silicon-based and non-conventional processors;

- Design mechanisms where the implementation of operating systems and middleware tasks alternates between software and hardware, so as to minimize and control power consumption and dissipation in a context of varying workload;

- Design compiling techniques that can enable compilers to automatically map fragments of code onto highly optimized hardware modules, like audio and video engines;

- Combine efficiency in power consumption with high performance requirements in parallel heterogeneous systems;

- Design new languages that efficiently exploit parallelism for massively heterogeneous and parallel systems, subject to strict power consumption constraints;

- ~~Develop scalable software for multilevel parallel heterogeneous systems, and which takes advantage of the architecture to ensure the desired concurrency level;~~

- Integrate multiple parallelism levels: *on-chip* (*multi-core*), *multi-thread*, and inter-node (*clusters, grids*);

- Provide support to the development of scalable computational systems;

- Quest into new computational models, as well as new architectures and machines to implement them;

- Investigate the possibility and perspectives of using biological mechanisms to solve problems;

- Create languages and methods to develop programs in machines that implement non-conventional computing models.

5. Participative and universal access to knowledge for the Brazilian citizen.

The information era has brought a revolution to the way people communicate and work. Computer networks allow people to share a diversity of resources (hardware, data and software, but also sound and visual information), regardless of their location or of actual physical presence. These new kinds of interactions are helped by the availability of wide band networks, with reduced latency, associated with mobile devices and ubiquitous computing. Availability, however, is not a synonym for ease of use or universal access.

Educational, technological, cultural, social and economical barriers hamper access to technology and interaction possibilities. The goal of this challenge is to breach these barriers by conceiving and implementing systems, tools, models, methods, procedures and theories able to address the issue of providing Brazilian citizens with access to knowledge. This access must be universal and participative, insofar as these citizens are not merely passive users who are at the receiving end of this knowledge, but actively participate in its generation and transformation. It is by allowing people to participate in the construction of knowledge that they will be able to fully and consciously make good use of this knowledge.

This problem concerns, therefore, extending computational systems to all Brazilians, respecting their diversity and differences. This challenge gains new proportions in the Brazilian scenario, where there are large socio-economic, cultural and regional differences, as well as unequal opportunities of access to technology and knowledge. Moreover, we are living a moment of media convergence (e.g., Internet, TV, cell phones) whose bottleneck is, undoubtedly, the possibility of universal access.

The 2003 census conducted by the IBGE (the Brazilian Institute of Geography and Statistics - Instituto Brasileiro de Geografia e Estatística) reported 32.1 million functional illiterates, defined as those over 15 years old with less than four years of schooling (26% of the population). Moreover, according to the same source, 24.5 million people (14.5% of the population) have some kind of physical handicap. Government, universities and private companies should search for technological solutions, of social impact, to decrease these differences and instill citizenship values in our society.

To approach this problem, it is necessary to define the terms "participative access" and "knowledge" in an encompassing way. It is not just a matter of capturing, organizing and disseminating information and knowledge. Neither is it a question of reducing the difficulties of physical access to computers and networks, whether caused by financial limitations, or by educational or cultural barriers, or by sensory or motor handicaps. This also involves producing computational technology that motivates and supports user participation in the process of knowledge production. Moreover, one must take into account juridical, social and anthropological aspects of Brazilian society in order to reduce the risks of aggravating such problems and to avoid generating new ones, augmenting the "digital exclusion" phenomenon.

The problem is difficult because it is unique, since there are no previous experiences to inspire solutions. This challenge demands multidisciplinary competences, in which Computer Science research will lead to creating systems and methods to sustain a digital culture for universal access to knowledge, while taking individual diversity and differences into account. The need to establish cooperation with other domains, such as Social and Human sciences, whose scientific practices differ from those of Computer Science, further complicates the issue

Several fields in Computer Science must contribute to the solution of this problem. Access to knowledge begins by conquering challenges in interfaces, in particular the interfaces between computing systems and their users. Other related open research topics involve the design and development of context-aware environments, using low cost hardware and software that is open and adaptable to local needs. The production of knowledge involves content production and efficient means to store and retrieve this content, thus impacting research on digital libraries. Affective, cultural and social issues in computing are new areas that must be involved in this task.

This challenge is characterized by an emphasis in interface design. The design for all and flexible and adjustable interfaces are state of the art research topics in the area of Human-Computer Interaction. Multimodal interaction plays a major role here – involving body, gaze, audio, speech or gesture, as well as the use of all kinds of sensors. The complexity of these issues increases as one considers the many socio-economic contexts in Brazil and the cultural diversity found in the country. Psychology, linguistics, anthropology, sociology and geography are among the areas needed to conquer the obstacles in interface research.

Concrete examples of applications that could benefit from results in this challenge include, among others, electronic government systems, learning systems, communities of practice or support to networked communities in underserved areas. It must be stressed that the term electronic government is not restricted to its more popular connotation – that of offering services on the Internet. Here, e-government means empowering citizens to generate knowledge to be shared, and to discuss subjects of their interest.

Research benefits involve the entire population and can reduce their distance to the greatest asset of today's society – knowledge. The Computer Science research community will also profit from this research, with new instruments, artifacts and methods for developing systems and user-friendly interfaces. Other scientific domains involved in this challenge will also gain by cooperation with computer scientists, by a better comprehension of the problem and of solution limitations. Still other benefits include new processes of knowledge socialization and the education of agents.

Besides work in flexible, multimodal and adaptable interfaces, some other research problems to attack in this challenge include:

- Design and development of new hardware and communications infrastructure;

- Experimentation with human-centered content, and content modeling, taking into account social dynamics and socially aware systems;

- Creation of *back-office* systems – the internal infrastructure needed to provide services to citizens, which can include long-term processes, involving several entities and thus issues of interoperability;

- Creation of the necessary infrastructure for allowing the citizen to interact with the processes conducted in the *back-office;*

- Development of supple mechanisms and structures for retrieval and storage of the content continuously generated by the user communities;

- Design and implementation of flexible and extensible ontological structures, to allow interoperability across knowledge domains and interactions among people of different vocabularies and cultural practices;

- Creation of *e-learning* platforms to allow efficient integration of communication tools to be used in electronic learning;

- Definition of means to ensure the appropriate administration of copyrights and of intellectual property in general, so as to allow a wide variety of experiments in knowledge production, administration and use; and

- Design and construction of new devices to support universal accessibility – e.g., helping physically handicapped users to fully interact with software and hardware systems.

6. Technological development of quality: dependable, scalable and ubiquitous systems

Information Technology is increasingly present in our daily lives. We do not need to go very far looking for examples. For instance, most home appliances and elevators are controlled via software; cars, tractors, planes, cell phones, systems to control urban traffic or surgery theaters also depend on this technology. While some of these examples correspond to relatively simple systems – a microwave oven – others are expensive and involve million of lines of code and sophisticated hardware. If this ubiquity brings comfort, it also causes problems. Since we depend on these systems, they must always be available and fault free; they must function as specified and be scalable and safe. This challenge thus concerns research in environments, methods, techniques, models, devices, design standards and architectures that can help designers and developers of large software and hardware systems to attain these goals.

The very many desirable properties covered by this challenge were grouped under a new term especially coined at the workshop to designate all these qualities – *omnivalent computing*. This term covers all concepts that relate ubiquity, dependability and evolution of all kinds of computational systems, especially software. Each of these properties presents research challenges: thus, their combination is a grand challenge.

Computer networks and the Internet have propelled the ubiquity of computational systems, It has acquired strength with the evolution in sensor technology – e.g., capturing data on temperature, pressure, atmospheric variables, sounds, electromagnetic waves, but also the identity of a person entering a room or brainwave patterns of patients with neurological disorders. Artificial eyes, noses, ears and taste buds are being developed to serve countless applications. Sensor devices are often considered to be a new class of computer systems, which differ from "normal" hardware by their ubiquity and collective analytical capacity. Sensors can vary in size and cost and range from those installed in satellites down to microssensors used, for instance, to identify pieces of clothing in department stores (*RFID*). The so-called actuator devices are complementary entities that, given data collected by a sensor, act on the environment with some specific goal in mind. The combination of sensors, software and actuators, in infinite variations, has revolutionized our society. Recent forecasts indicate that, within the next decade, sensors and distributed computing will invade all environments - homes, offices, factories, cars, streets, or farms.

From the application viewpoint, wireless sensor networks can be used in several scenarios, including environmental monitoring, agriculture, event tracing, coordination of actions, and information mining and processing. A Brazilian example is the combustion system of Flex (dual combustion) cars, automatically regulated by a sensor in the exhaust pipe, which determines the composition of the gases produced, and provides this feedback to the engine.

The proliferation of types and uses of such devices has prompted increasing attention to the development of systems to coordinate them. Indeed, once put in place, sensors should ideally be always available. Their networks should be scalable (functioning as expected regardless of their expansion) and safe (preventing malicious attacks). Finally, sensors should work correctly, i.e., according to their specification. Ensuring all these characteristics is still an open problem.

Sensor networks were used to illustrate some of the issues that motivate this challenge – they are ubiquitous and present a combination of hardware and problems that threatens their dependability. While sensor network technology is relatively new, software systems themselves are also ubiquitous and suffer from the same dependability problems. Unlike sensor networks, however, software problems have been present for several decades. Software dependability and scalability are problems that occur both in life-critical applications (e.g., those that control airplanes, nuclear plants or orbital stations) as well as non-critical ones (e.g., payroll or warehouse systems, where such problems, though serious, can be eventually repaired without endangering lives). The need to develop software that meets all these requirements is causing a revolution in Software Engineering and associated disciplines.

The construction and maintenance of robust software demand qualified professionals and long-term commitments. Reliable software engineering means far more than incorporating fault tolerance and ensuring correctness. The notion of dependability varies with the service the software is expected to perform. In some cases, for instance, applications involving command and control, an error may have catastrophic effects. In other cases, e.g. Web information retrieval, an incorrect answer (false positive or absence of relevant answers) is acceptable, as long as this does not happen frequently. A system that cancels its own execution, invades the user's privacy or presents a risk to whoever uses it is just not acceptable.

Given the growing presence of software in everyone's lives, ensuring its dependability becomes increasingly critical. Recent research concerning software quality problems points out that around 50% of the software that is made available contains non-trivial errors. Another issue is software evolution, an intrinsic property, i.e., all software tends to evolve. Supporting this evolution makes up a considerable percentage of its total cost (*Total Cost of Ownership*). This in turn requires the need for developing software that can evolve without compromising its quality. Finally, it is a known fact that, regardless of how rigorously a software program is developed, it will contain errors. Thus, one must live with failure – of use, of hardware or of software - while at the same time preventing that errors cause an unacceptable damage level.

Brazil, like India and Ireland, aspires to a powerful software industry. In spite of all efforts in this direction, results are still not impressive. Improving software quality will certainly contribute to its acceptance, in Brazil and abroad. This is also the best way to ensure growth in the internal market.

These problems in producing software are hard when one "just" considers software with traditional architectures. This tends to become worse in the scenario of geographically distributed software development (several teams developing a piece of software) and of distributed software operation (several CPUs with distinct modules of the same software, serving the same application, such as in *grid* environments). Another factor that makes dependable and ubiquitous development a challenge is the growing complexity in software (both in increase in volume and in the scope and constraints on quality requirements). Moreover, demand for 24/7 availability and proliferation of non-expert users are additional complicating factors.

Many are the benefits that can result from attacking this challenge. Several of these result from the advantages of implementing reliable and secure systems, thus contributing to quality improvement and the reduction of costs in development and use of computational systems. Computational system malfunctions can have serious consequences. For instance, failures in systems that control high risk processes can result in the loss of lives, ecological disasters, the crashing of companies. In electronic commerce systems, errors can lead to millions of dollars in losses. In electronic government, they can contribute to social exclusion.

Errors in security of data, software or systems also result in social and financial losses, such as those caused by network invasions or viruses all over the world. Privacy violation resulting from failures is another serious consequence of such problems, with unimaginable costs to all citizens.

Given these remarks, some of the topics associated with this challenge include:

- Development and evaluation of models and modeling tools for software systems that rely on a solid theoretical basis;

- Development and adaptation of technologies and all kinds of tools to support implementation and evaluation of software that is dependable by construction;

- Development of tools to support the process of software implementation and evolution;

- Specification and analysis of new algorithms and techniques for data and system security, including cryptography and secure communication protocols; and

- Construction of mechanisms and tools to support fault tolerance, permanent availability and scalability;

- Considering the need for ubiquity in systems design and development, including distinct work environments and varying requirements.

7. Some proposed actions

The workshop's participants proposed a few actions associated with the five challenges discussed. These are just examples of a great number of other actions that can be undertaken. The main issues discussed involved the following items:

<u>Multidisciplinarity</u> – the evolution of research and development in the 21st century points at multidisciplinary groups a factor needed to succeed in obtaining scientific results. Thus, two actions were recommended in this item: (a) to make Computer Science researchers aware of the problems inherent to multidisciplinary research, such as establishing common vocabularies and understanding the differences among research methodologies in distinct knowledge domains; and (b) to develop "*joint venture*" research and development models across areas, to educate professionals and scientists to work in this new world, emphasizing multi- and inter-disciplinarity. Examples of multidisciplinary applications that can motivate this kind of learning are those of public health, environmental problems, urban violence, agriculture, *e-learning*, digital entertainment, telemedicine or history, among others.

Multidisciplinarity should be fostered not only between Computer Science and other scientific domains, but also within Computer Science itself. For instance, experts in hardware need to cooperate with researchers in networks, databases, and human-machine interaction. All of these people must themselves interact continuously with researchers in computer graphics, scientific visualization, artificial intelligence and other areas necessary to solve the challenges.

<u>Integration with industry</u> – good quality research brings about social and economic benefits. Thus, another action considered was to interact and get closer to the industrial sector to ensure technological development with quality and to detect new areas with the potential to give origin to emerging markets.

<u>Transform brain drain into an asset</u> – many Brazilian researchers are being attracted by more favorable work conditions abroad. These people can be taken advantage of as contacts to increase cooperation in research between Brazil and the rest of the world – e.g., by helping to promote international conferences in Brazil, or bringing visiting scholars.

<u>Founding a center for research brainstorming</u> – workshop participants were unanimous to remark that "brainstorming" events, such as this workshop, are extremely rare. Most conferences and seminars all over the world follow a standard pattern of presentation of papers and discussion of specific themes. The action proposed is to found a center for scientific meetings, similar to Germany's Dagstuhl or Oberwolfach. The first was created to provide computer scientists with conditions to develop leading edge research. One of the best-known Dagstuhl activities, for instance, are its one-week Computer Science international seminars, for at most 40 participants, to discuss and to develop top research. Each such conference is proposed up to three years in advance to a scientific committee, and participants –pay only transportation costs to and from the center. Oberwolfach installations serve a similar purpose for researchers on Mathematics and Theory of Computing (e.g., see <u>www.dagstuhl.de</u>).

8. Conclusions

The Brazilian Computer Society hopes that this is the first of a series of meetings to discuss Grand Challenges of Computer Science in Brazil, having in mind long-term planning of CS research in the country.

As mentioned in the Introduction, the idea of formulating Grand Research Challenges has been adopted in several countries, in distinct fields. In these countries, it has been observed that the formulation of these challenges has led to establishing a long-term research agenda, with positive consequences not only in terms of advancing knowledge, but also to educate new generations of researchers. Thus, this proposal from the Society can bring significant contributions to Computer Science in Brazil and also help the formulation of research policies of funding agencies.

The proposed challenges require cooperation within Computer Science and with other fields. Each challenge involves extensive innovative research, being centered on some important aspects of computing. Several research areas in Computer Science appeared in the description of all challenges – computer networks, data management, software development, design of algorithms and data structures, interface design, and new devices and computer architectures are some of the issues to be treated with less or more detail in all items.

Finally, Computer Science permeates all scientific activity nowadays. SBC hopes, thus, that this initiative will also contribute to advances in Brazilian science on all fronts, influencing many other research fields in the country. Reports produced all over the world stress the fundamental role of Computer Science in the technological, scientific, economic and social progress of a nation. The search for scientific excellence in Computer Science, with emphasis in a long-term view, should therefore have lasting social and economical impact.

Bibliographic references

[1] Computational Science: Ensuring America´s Competitiveness. PITAC Report to the President, USA, June 2005, Available in http://www.nitrd.gov/pitac/reports/20050609_computational/computational.pdf

[2] .G. Câmara. Grand Challenges in Computer Science: The construction of a third culture. Presentation for the Grand Challenges workshop, may 2006

Workshop program

Day 1:

8:30-12:30 Presentation of the selected position papers – 8 minutes per proposal
14:00-16:00 Panel – Renato Janine Ribeiro and Gilberto Camara – The importance of long-term planning of Computer Science research in Brazil
16:30 – 19:30 Organization of 5 working groups, to discuss and consolidate the Grand Challenge proposals. These grand challenges were derived from a consensual view of each group on future research issues.
Presentation of the work of each group in posters.

Day 2:

8:30-12:30 Group presentations: summing up and consolidation of the Grand Challenges proposed in the previous evening
14:00-15:00 Discussion and final consolidation of the challenges, with justification
15:00- 17:00 Final presentation of the Grand Challenges in Computer Science in Brazil 2006-2016 to invited guests.