Cyber-Physical Systems: A New Frontier

Lui Sha[1], Sathish Gopalakrishnan[2], Xue Liu[3], and Qixin Wang[1]
[1]University of Illinois at Urbana Champaign, [2]University of British Columbia, [3]McGill University
lrs@cs.uiuc.edu; sathish@ece.ubc.ca; xueliu@cs.mcgill.ca; qwang4@uiuc.edu

Abstract: The report of the President's Council of Advisors on Science and Technology (PCAST) has placed CPS on the top of the priority list for federal research investment [6]. This article first reviews some of the challenges and promises of CPS, followed by an articulation of some specific challenges and promises that are more closely related to the Sensor Networks, Ubiquitous and Trustworthy Computing Conference.

1. Introduction

The Internet has made the world "flat" by transcending space. We can now interact with people and get useful information around the globe in a fraction of a second. The Internet has transformed how we conduct research, studies, business, services, and entertainment. However, there is still a serious gap between the cyber world, where information is exchanged and transformed, and the physical world in which we live. The emerging cyber-physical systems shall enable a modern grand vision for societal-level services that transcend space and time at scales never possible before.

Two of the greatest challenges of our time are global warming coupled with energy shortage, and the rapid aging of a significant fraction of the world's population with the related chronic diseases that threaten to bankrupt healthcare services, such as Medicare, or to dramatically cut back medical benefits.

During the meeting of the World Business Council for Sustainable Development in Beijing on March 29, 2006, George David[1] noted: *"More than 90 percent of the energy coming out of the ground is wasted and doesn't end as useful. This is the measure of what's in front of us and why we should be excited."* Buildings and transportation are sectors with heavy energy consumption. During the NSF CDI Symposium (September 5-6, 2007) at RPI, Clas A. Jacobson[2] noted that green buildings hold great promises. Energy used in lightening and cooling buildings is estimated at 3.3 trillion KWh. Technologically, it is possible to reach the state of Net Zero Energy Buildings, where 60-70% efficiency gains required for reducing demand and balance to be supplied by renewable. However, to reach the goal of net zero energy buildings, we must tightly integrate the cyber world and the physical world. He noted that in the past the science of computation has systematically abstracted away the physical world and vice versa. It is time to construct a Hybrid Systems Science that is simultaneously computational and physical, providing us with a unified framework for robust design flow with multi-scale dynamics and with integrated wired and wireless networking for managing the flows of mass, energy, and information in a coherent way.

According to the Department of Energy, the transportation share of the United States' energy use reached 28.4% in 2006, which is the highest share recorded since 1970[3]. In the United States, passenger and cargo airline operations alone required 19.6 billion gallons of jet fuel in 2006. According to Time[4], 88% of all trips in the U.S. are by car. Work related needs including daily work commute and business travel is a significant fraction of the transportation cost. Telepresence research seeks to make all

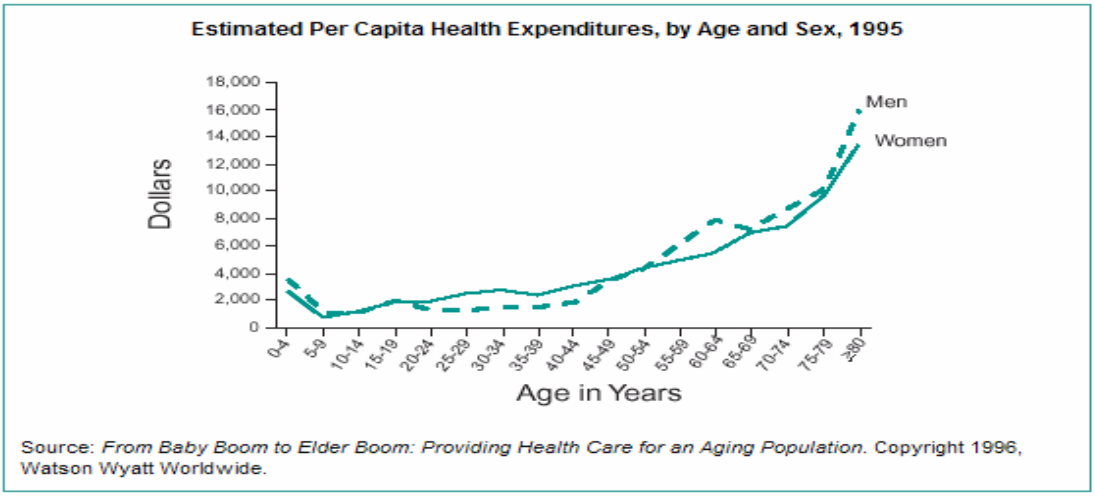
[1] Chairman and CEO of United Technology Research Center

[2] Chief Scientist, Control, United Technology Research Center

[3] http://cta.ornl.gov/data/new_for_edition26.shtml

[4] www.time.com/time/specials/2007/environment/article/0,28804,1602354_1603074_1603122,00.html

978-0-7695-3158-8/08 $25.00 © 2008 IEEE
DOI 10.1109/SUTC.2008.85



Source: *From Baby Boom to Elder Boom: Providing Health Care for an Aging Population*. Copyright 1996, Watson Wyatt Worldwide.

interactions seem local rather than remote. It is one of the three grand challenges of in multimedia research[5] to make interactions with remote people and environments nearly the same as interactions with local people and environments. Integrating wired and wireless networks with real-time, interactive, immersive three-dimensional environments and tele-operation can minimize work-related travel.

The rapidly aging population with age related chronic diseases is another formidable societal challenge. It is alarming to note that the growth of per-capita health cost has been increasing near exponentially with an increase in the age of the population.

According to the CDC[6], more than 90 million Americans live with chronic illnesses.

- Chronic diseases account for 70% of all deaths in the United States.
- The medical care costs of people with chronic diseases account for more than 75% of the nation's $1.4 trillion medical care costs.
- Chronic diseases account for one-third of the years of potential life lost before age 65.

Advanced biotechnology holds great promise to improve the heath of an aging population. For example, stem-cell biotechnology holds the promise of treatment for many age-related diseases. According to NIH [7], "stem cells, directed to differentiate into specific cell types, offer the possibility of a renewable source of replacement cells and tissues to treat diseases including Parkinson's and Alzheimer's diseases, spinal cord injury, stroke, burns, heart disease, diabetes, osteoarthritis, and rheumatoid arthritis." In addition, "Human stem cells could also be used to test new drugs. For example, new medications could be tested for safety on differentiated cells generated from human pluripotent cell lines."

However, much of this potential is not tapped, largely due to lack of sufficient knowledge of the complex and dynamic stem-cell microenvironment, also known as the niche. There is a need to mimic niche conditions precisely in artificial environments to correctly regulate stem cells ex vivo. Indeed, the sensing and the control of the stem cell microenvironment are at the frontier of stem cell research. According to Badri Roysam[8] the stem cell niche has a complex multi-cellular architecture that has many parameters, including multiple cell types related by lineage, preferred

[5] http://delivery.acm.org/10.1145/1050000/1047938/p3-rowe.pdf?key1=1047938&key2=8175939811&coll=GUIDE&dl=GUIDE&CFID=15151515&CFTOKEN=6184618

[6] http://www.cdc.gov/nccdphp/overview.htm#2

[7] http://stemcells.nih.gov/info/basics/basics6.asp

[8] Professor, ECSE & Biomedical Engineering, RPI and Associate Director, NSF ERC Center for Subsurface Sensing & Imaging Systems

spatial locations and orientations of cells relative to blood vessels, soluble factors, insoluble factors related to the extra-cellular matrix, bio-electrical factors, biomechanical factors, and geometrical factors. The combinatorial space of parameter optimization and niche environment control calls is a grand challenge in embedded sensing and actuation.

A closely related problem is providing care to the elderly population without sending them to expensive nursing homes. In the United States alone, the number of people over age 65 is expected to reach 70 million by 2030, doubling from 35 million in 2000. Expenditure in the United States for health-care will grow to 15.9% of the GDP ($2.6 trillion) by 2010. Unless the cost of health care for the elderly can be significantly reduced, financially stressed Social Security and Medicare/Medicaid systems will likely lead to burdensome tax increases and/or benefit reductions. A major cost is the loss of the ability to remain in the home because of the need for greater health care supervision.

One crucial factor contributing to the loss of independence and the resulting institutionalization is the need for assistance in physical mobility. Another key factor is cognitive impairment that requires daily supervision of medication and health-condition monitoring. When future CPS infrastructure supports tele-presence, persons with one or more minor mobility impairments can regain their freedom of movement at home. In addition, physiological parameters critical to the medical maintenance of health can be monitored remotely. When the elderly can maintain their independent living without loss of privacy, a major financial saving in senior care will result. Furthermore, the elderly will be much happier by living independently at home while staying in contact with their social networks.

As observed by SUTC 2008, "*rapid research and technological advances in wireless communications and increasing availability of sensors, actuators, and mobile devices have created an exciting new ubiquitous computing paradigm that facilitates computing and communication services all the time, everywhere. This emerging paradigm is changing the way we live and work today. Via a ubiquitous infrastructure consisting of a variety of global and localized networks, users, sensors, devices, systems and applications may seamlessly interact with each other and even the physical world in unprecedented ways. To realize this continually evolving ubiquitous computing paradigm, trustworthy computing that delivers secure, private, and reliable computing and communication services play an essential role*."

Clearly, SUTC research has a key role to play in the emerging cyber-physical system of systems. A variety of questions need to be answered, at different layers of the architecture and from different aspects of systems design, to trigger and to ease the integration of the physical and cyber worlds.

2. The Challenges of Cyber-Physical System Research

2.1. Real-time System Abstractions

Future distributed sensors, actuators, and mobile devices with both deterministic and stochastic data traffic require a new paradigm for real-time resource management that goes far beyond traditional methods. The interconnection topology of mobile devices is dynamic and the system infrastructure can also be dynamically reconfigured in order to contain system disruptions or optimize system performance. There is a need for novel distributed real-time computing and real-time group communication methods for dynamic topology control in wireless CPS systems with mobile components with dynamic topology control. Understanding and eventually controlling the impact of reconfigurable topologies on real-time performance, safety, security, and robustness will have tremendous impact in distributed CPS system architecture design and control.

Existing hardware design and programming abstractions for computing are largely built on the premise that the principal task of a computer is data transformation. Yet cyber-physical systems are real-time systems. This requires a critical re-examination of existing hardware and software architectures that have been built over the last several decades. There are foundational opportunities that have the potential of defining the landscape of computation in the cyber-physical world. When computation interacts with the physical world, we need to explicitly deal with events distributed in space and time. Timing and spatial information need to be explicitly captured into programming models. Other physical and logical properties such as physical laws, safety, or power constraints, resources, robustness, and security characteristics should be captured in a composable manner in programming abstractions. Such programming abstractions may necessitate a dramatic rethinking of the traditional split between programming languages and operating systems. Similar changes are required at the software/hardware level given performance, flexibility, and power tradeoffs.

We also need strong real-time concurrent programming abstractions. Such abstractions should be built upon a model of simultaneity: bands in which the system delays are much smaller than the time constant of the physical phenomenon of interest. The programming abstractions that are needed should also capture the ability of software artifacts to execute at multiple capability levels. This is motivated by the need for software components to migrate within a cyber-physical system and execute on devices with different capabilities. Software designers should be capable of expressing the deprecated functionality of a software component when it executes on a device with limited resources.

The programming abstractions that we envision will need support at the middleware and operating system layers for:

- Real-time event triggers,
- Consistent views of distributed states in real-time within the sphere of influence. This challenge is especially great in mobile devices
- Topology control and "dynamic real-time groups" in the form of packaged service classes of bounded delay, jitter and loss under precisely specified conditions,
- Interface to access to the same type of controls regardless of the underlying network technology.

2.2. Robustness, Safety and Security of Cyber-Physical Systems

Uncertainty in the environment, security attacks, and errors in physical devices and in wireless communication pose a critical challenge to ensure overall system robustness, security and safety. Unfortunately, it is also one of the least understood challenges in cyber-physical systems. There is a clear intellectual opportunity in laying the scientific foundations for robustness, security and safety of cyber-physical systems in general and in SUTC systems in particular. An immediate aim should be to establish a prototypical SUTC model challenge problems and to establish a set of useful and coherent metrics that capture uncertainty, errors, faults, failures and security attacks.

We have long accepted that perfect physical devices are rare. A perfect example of this approach is the design of reliable communication protocols that use an inherently error-prone medium, whether wired or wireless. This prudence has not been applied to other software engineering processes. We have depended, more often than not, on the correctness of the results from our microprocessors and other hardware elements. While we have successfully masked many hardware failures using a combination of innovative circuit design, redundancy and replay, we have largely regarded most other errors as either transient – caused by bit flips – or permanent. Transient errors can be ameliorated by re-execution and permanent failures require migrating tasks to fault-free hardware. Sub-micron scaling of

semiconductor devices and device density, however, will present us with hardware that is more error-prone, and errors are likely to be neither transient nor permanent. Intermittent errors – that last several milliseconds to a few seconds – may not be uncommon in future generation chip multiprocessors [9]. To tolerate intermittent failures, we will likely need to apply algorithms that do not rely on the accuracy of one computation. Ideas concerning imprecise computations [11] will gain more relevance; developing algorithms using those principles will be extremely valuable on the road to robust systems.

These trends will, however, make our current efforts of build perfect software more difficult. Indeed, there has been great advancement in automated theorem proving and model checking in recent years. However, it is important to remember that cyber-physical systems are real-time systems and the complexity of verifying temporal logic specifications is exponential. That is, like the physical counterpart, a perfect software component is also rare and will remain that way. This has profound implications. We need to invent a cyber-physical system architecture in which the safety critical services of large and complex CPS can be guaranteed by a small subset of modules and their interactions; the design of this subset will have to be formally specified and verified. Their assumptions about the physical environments should be fully tested, and furthermore, we need to develop advanced and integrated static analysis and testing technologies to ensure that 1) the software code is compliant with the design, and that 2) the assumptions regarding external environment are sound. Finally, cyber-physical systems are deeply embedded and they will evolve in place. The verification and validation of cyber-physical system is not a one-time event; it should be life cycle process that produces an explicit body of evidence for the certification of safety critical services.

Safety critical services apart, we still have the great challenge of how to handle known and unknown residual errors, and security gaps in many useful, but not safety critical, cyber-physical components that have not been fully verified and validated. The broadcast nature of a wireless network and interference make these challenges more serious. In physical systems, it is the theory of feedback control that provides the very foundation to achieve robustness and stability despite uncertainty in the environment and errors in sensing and control. The current open loop architecture in software systems may allow a minor error to cascade into system failure. The loops must be closed across both the cyber world and physical world. The system must have the capability to effectively counter-act uncertainties, faults, failures and attacks. The recent development of formal specification based automatic generation of system behavior monitoring, the steering of computation trajectories, and the use of analytically redundant modules based on different principles, while still in infancy, is an encouraging development.

Safety, robustness and security of the composed CPS also require explicit and machine checkable assumptions regarding external environments; formally specified and verifiable reduced complexity critical services and reduced complexity interaction involving safety critical and non-safety critical components; and analytically redundant sensing and control subsystems based on different physical principles and/or algorithms so as to avoid common mode failures due to faults or attacks. We also need theory and tools to design and ensure well-formed dependency relations between components with different criticality as they share resources and interact. Stable and robust upgrading of running systems will be another critical aspect of cyber-physical systems, especially in large critical infrastructure systems that cannot or too expensive to shut down.

2.3. System QoS Composition Challenge

CPS systems are distributed and hybrid real-time dynamic systems, with many loops of different degree of application criticality operating at different time and space scales. Compositional system modeling, analysis, synthesis and

integration for such systems are at the frontier of research. The "science" of system composition has clearly emerged as one of the grand themes driving many of our research questions in networking and distributed systems. By *system composition* we mean that the QoS properties and functional correctness of the system can be derived from the architectural structure, subsystem interaction protocols, and the local QoS properties and functional properties of various constituent components.

A framework for system composition should highlight the manner of the composition of components and the methods to derive QoS metrics for a composite system.

Current compositional frameworks have been developed with limited heterogeneity, such as real-time resource management, automata and differential equations. The new theory of system composition must provide a comprehensive treatment of system integration concerns. Each component should provide a system composition interface that specifies not only its input and output but also relevant QoS properties and constraints. In electronic subsystems, the properties of a circuit depend not only on component properties but also on how they are connected together. Likewise, a CPS system's properties will depend on both component properties and the structure of system architecture.

The framework for capturing subsystem requirements needs to be powerful enough to describe both deterministic requirements and probabilistic requirements. Not all subsystems in a CPS will be hard real-time systems and methods need to evolve to capture different types of requirements, to synthesize compositional requirements, and to determine the feasibility of meeting those requirements.

Large CPS systems will have many different QoS properties encompassing stability, robustness, schedulability, security, each of which will employ a different set of protocols and will need to be analyzed using a different theory. It is important to note that these protocols may not be orthogonal and, sometimes, could have pathological interactions; for example, the well-known problem of unbounded priority inversion when we use synchronization protocols and real-time priority assignments as is. There are also numerous reports from the field about the adverse interactions between certain security, real-time and fault tolerant protocols. Thus, the theory of system composition must address not only the composability at each QoS dimension but also the question of how the protocols interact.

2.4. Systems Engineering Research

From a systems engineering perspective, we need a scientific methodology to iteratively build both the system structure model and the system behavior model. We need to develop analytical capability to map behavior onto structure and vice versa so that we identify what aspects of the required behavior will be performed by which specific parts of structure. We need techniques to perform quantitative trade-off analysis that will take into account the available technology and constraints on the cyber components, from the physical components, and from human operators. In the development of CPS systems, constraints imposed by the physical sciences (such as physics, chemistry, materials science) will need to interact with constraints on the computing artifacts (such as computational complexity, robustness, safety and security). To make fundamental progress, we need a combination of model-based system and software design and integration technologies; and deep analysis of the underlying abstractions and their interactions.

With this scientific and engineering framework, we must be able to judiciously choose the location, computing and communication capabilities as well as energy reserves of network nodes in order to handle the required data flows efficiently. We should be able to alter the topological characteristics of the network by changing transmission power, medium access control, and communication protocols. Our challenge is to

formulate a new calculus that merges time-triggered and event-driven systems. We need it to be applicable to hierarchies that involve dynamics at drastically different time scales from months to microseconds and geographic scope from on chip to the world. This is a grand challenge.

2.5. Trust in Cyber-Physical Systems

Users of cyber-physical systems will need to place a high level of trust in the operation of the systems. Trust is a combination of a many characteristics, mainly *reliability, safety, security, privacy and usability*.

System models and abstractions (described earlier) have to incorporate fault models and recovery policies that reflect the scale, lifetime, distributed control and replace/reparability of components. Safety, as we have mentioned in brief in prior sections, requires attention in a larger context as well. The ubiquitous use of CPS applications should not limit the availability of alternative systems – social and technological – that can handle large-scale failures. We would need to create guards that ensure that the automation does not increase hazards when compared to the non-automated system. The extra emphasis on safety is to highlight the need for tools that provide support for visualizing and analyzing a cyber-physical within the broader context of other social and cyber-physical systems. Two cyber-physical systems may never interact directly but may be coupled by human behavior, and we need to understand the nature of such interactions and reason about safety as a global, not local, property.

An increased dependence on cyber-physical systems will lead to the collection of a vast amount of human-centric data at various scales. Although cyber-physical systems greatly enrich our life qualities and experiences, they also bring about privacy and security concerns [10]. For example, in applications such as assisted living [5] and wireless medical device networks (Section 3), private personal data and medical data should be protected with different levels of information disclosure to different roles (health care providers, medical team, relatives, or assisted persons). Unauthorized access to private information can have serious consequences. We need an analytical foundation and the associated engineering framework to address privacy protection in cyber-physical systems. A combination of mechanisms for auditing and regulating access to information, for preserving privacy of individuals but exporting aggregate statistics, and legal procedures for enforcement of privacy protection would need to evolve to make cyber-physical systems acceptable to a large population.

All systems become *usable* when complexity that does not need to be exposed to users is kept hidden, and when unavoidable complexity is exposed to users according to cohesive, conceptual models that maximizes system predictability, supports users' abilities to generalize about such behavior, and minimizes corner cases. Usability poses a variety of problems involving human cognition, computer-human interaction and interface design.

3. Medical Device Network: An Example Cyber-Physical System

An example on how to design medical device network may provide us better understanding of the aforementioned challenges. As noted in the report of the NSF High-Confidence Medical Device Software and Systems (HCMDSS) workshop [7]: "*Advances in computing, networking, sensing, and medical-device technology are enabling the dramatic proliferation of diagnostic and therapeutic devices. Those devices range from advanced imaging machines to minimally invasive surgical techniques, from camera-pills to doctor-on-a-chip, from computerized insulin pumps to implantable heart devices. Although advances in standalone diagnostic and treatment systems have been accelerating steadily, the lack of proper integration and interoperation of those systems produces systemic inefficiencies in health care*

delivery. This inflates costs and contributes to avoidable medical errors that degrade patient care. The use of software that controls medical devices to overcome these problems is inevitable and will ensure safe advances in health care delivery. The crucial issue, however, is the cost-effective development and production of reliable and safe medical-device software and systems."

The next generation medical system is envisioned as a ubiquitous system of wired and wireless networked medical devices and medical information systems for secured, reliable, privacy-preserving health care. It will be a networked system that improves the quality of life. For example, during a surgical operation, context information such as sensitivity to certain drugs will be automatically routed to relevant devices (such as infusion pumps) to support personalized care and safety management. A patient's reactions – changes in vital signs – to medication and surgical procedures will be correlated with streams of imaging data; streams will be selected and displayed, in the appropriate format and in real time, to medical personnel according to their needs, e.g., surgeons, nurses, anesthetists and anesthesiologists. During particularly difficult stages of a rare surgical operation, an expert surgeon can remotely carry out key steps using remote displays and robot-assisted surgical machines, sparing the surgeon of the need to fly across the country to perform, say, a fifteen-minute procedure. Furthermore, data recording will be integrated with storage management such that surgeons can review operations and key findings for longitudinal studies for the efficacy of drugs and operational procedures.

While networked medical devices hold many promises, they also raise many challenges. First, from operating rooms to enterprise systems, different devices and subnets have different levels of clinical criticality. Data streams with different time sensitivities and criticality levels may share many hardware and software infrastructure resources. How to maintain safety in an integrated system is a major challenge that consists of many research issues. Indeed, many medical devices are safety critical and must be certified. Thus, it is important to develop a standard-based, certifiable wire and wireless networked medical devices infrastructure to lower the cost of development, approval, and deployment of new technologies/devices. The development of technologies that can formally specify both the application context and the device behaviors is a major challenge for the vision of certifiable plug and play medical devices in the future.

Second, most monitoring devices are being moved from wired networks to wireless networks. How do we provide on-demand reliable real-time streaming of critical medical information in a wireless network? This is another hard problem. For example, when an EKG device detects potentially dangerous and abnormal heartbeats, it is of critical importance to ensure that not only the warning but also the real-time EKG streams are reliably displayed at nursing stations. Furthermore, reliable on demand real-time streaming must coexist with other wireless devices. For example, in an intensive care unit, we have 802.11 wireless networks, cellular phones, wireless PDAs, RFID, two-way radios and other RF emitting devices. This necessitates a network infrastructure to reliably integrate myriad wireless devices, to let them coexist safely, reliably and efficiently. To address these concerns, the FDA has issued an official guideline for medical wireless network development [3].

To design an integrated wired and wireless medical device network, we face all the aforementioned QoS composition challenges. For example, how does one monitor and enforce safe, secure, reliable and real time sharing of various resources, in particular the wireless spectrum? How does one balance the resources dedicated to reliability, real-time performance and the need for coexistence? What is the programming paradigm and system composition architecture to support safe and secured medical device plug and play [8]?

Acknowledgements. Most of the material presented here originated from discussions, presentations, and working group documents from NSF workshops on Real-time GENI and from

NSF workshops on Cyber-Physical Systems [1][2]. The authors thank all the workshop participants for their insightful contributions.

References

[1] Real-time GENI report. http://www.geni.net/GDD/GDD-06-32.pdf

[2] NSF Workshops on Cyber Physical Systems. http://varma.ece.cmu.edu/cps/

[3] FDA, Draft Guidance for Industry and FDA Staff – Radio-Frequency Wireless Technology in Medical Devices, Jan. 2007. http://www.fda.gov:80/cdrh/osel/guidance/1618.html

[4] Mu Sun, Qixin Wang, and Lui Sha, "Building Reliable MD PnP Systems", Proceedings of the Joint Workshop on High Confidence Medical Devices, Software, and Systems and Medical Device Plug-and-Play Interoperability, Jun. 2007.

[5] Qixin Wang, et al., "I-Living: An open system architecture for assisted living," Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Oct. 2006, pp. 4268-4275.

[6] http://ostp.gov/pdf/nitrd_review.pdf

[7] Insup Lee, et al., High-confidence medical device software and systems. http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/2/33950/01620992.pdf

[8] http://www.mdpnp.org/Home_Page.html

[9] Phillip M. Wells, Koushik Chakraborty, and Gurindar S. Sohi, "Adapting to intermittent faults in future multicore systems," Proceedings of the International Conference on Parallel Architectures and Compilation Techniques, Sept. 2007.

[10] Jaideep Vaidya and Chris Clifton. "Privacy-preserving data mining", IEEE Security & Privacy Magazine, Vol. 2, No. 6, Nov.-Dec. 2004, pp. 19 – 26.

[11] Jane W.-S. Liu, et al., "Imprecise computations", Proceedings of the IEEE, Vol. 82, No. 1, Jan. 1994, pp. 83 – 94.