Real Time and Embedded (RTE) GENI

Lui Sha

University of Illinois at Urbana Champaign

Ashok Agrawala

University of Maryland at College Park

Foreword: This is a brief pre-view of the NSF Real Time and Embedded workshop report. Any opinion, finding, conclusion or recommendation expressed in this article are those of the authors and do not necessarily reflect the views of NSF and may differ from the final report on RTE GENI.

1.0 Introduction

Looking back at the big picture of computing, the convergence of computing and networking has given us the Internet. This merging of technologies has impacted all aspects of society: it has fundamentally changed how corporations, governments, and ordinary citizens receive, manage, and deliver information and services. The basic architecture of the Internet was developed during the 1970s, and has proven remarkably adaptive over the decades with all aspects of technology, computation, storage and communication improving by several orders of magnitude. More importantly, the basic architecture has successfully supported a wide variety of applications, including interactive applications, global-scale information services, and bulk-data transfer applications. Higher layer protocols for strong security have also been successfully layered, leading to a comprehensive suite of applications for digital commerce, secure mail, anonymous communications, and so on.

However, there is a mismatch between the design of the original "best-effort" Internet protocols and the requirements of the real-time (RT) applications that are delay and jitter sensitive. Real-time control applications consist primarily of connection-oriented feedback loops, sensor data feeds, and actuation signal streams. The performance and stability of these applications is sensitive to loss, delay, and jitter. This mismatch between Internet architecture and fundamental requirements of real-time applications forced the RT industry to create a number of *one-off* domain specific protocols. These include:

- Factory automation and process control protocols such as PROFIBUS[1], and the recent Field Bus standards[2] .
- Building automation protocols such as BACnet[3], EIB/KNX[4], LonWorks[5].
- Protocols in the automotive and aircraft domain, e.g. CAN, FlexRay[7], TTP[8], and ARINC[9].

The proliferation of such special-purpose real time networking standards increases cost and creates inter-operability problems when these networks are linked together. Furthermore, advances in sensors, wired and wireless networks, and low-power intelligent devices are enabling an entirely new class of networked applications in the sensing, monitoring, and control domain. There is also an overwhelming consensus emerging in the networking community that we need to rethink the Internet Architecture with a clean slate. This is prompted by:

- Fundamental architectural limitations of the current Internet and continued architectural erosion making things worse with time.
- Inability to incorporate emerging disruptive technologies and new applications.
- Inability to support many existing real time applications domains mentioned above.

The purpose of these workshops was to engage research and user communities to create the requirement profiles and the scientific and technological foundation to foster rapid convergence of computing, communication, intelligent sensing and control of our physical environment. This coming great wave, when it materializes, will profoundly change how we live.

In the next section, we will review several distributed real-time applications to motive the research, followed by a summary of the challenges to realize the vision of RTE GENI.

2.0 Applications for Real Time and Embedded GENI

RTE GENI is about building the infrastructure for systems of distributed real time and embedded systems. In this section, we will take a look at challenges raised by some of the existing and futuristic applications.

2.1 Medical Device Network and Health Management Systems

The next generation medical system is envisioned to be a ubiquitous system of networked systems for secure, reliable, privacy-preserving, cost-effective, and personalized quality health care leading to not only a better health-care delivery but improving the quality of life in general. Unfortunately, current medical devices are still mostly stand alone systems with proprietary designs. Medical workers often need to manually transfer or enter data between machines. They must mentally correlate many paper records and screen displays from different diagnostic and monitoring machines. Sometimes, they also need to manually set different devices in such a way to emulate the interlock needed between different devices and actions. This is time consuming, burdensome and error prone process. When a medical worker's attention is distracted, tragedies may occur. For example:

> *"A 32-year-old woman was having a laparoscopic cholecystectomy performed under general anesthesia. During that procedure and at the surgeon's request, a plain film x-ray was shot during a cholangiogram. The anesthesiologist stopped the ventilator for the x-ray . The x-ray technician was unable to remove the film because of its position beneath the table. The anesthesiologist attempted to help the technician, but found it difficult because the gears on the table had jammed. Finally, the x-ray was removed, and the surgical procedure recommenced. At some point, the anesthesiologist glanced at the EKG and noticed severe bradycardia. He realized he had never restarted the ventilator. This patient ultimately died."* APSF Newsletter, Winter 2005.

If safety interlock were installed, i.e., the ventilation machine could only be disabled for a short and safe duration when the X-Ray machine button is pushed, the tragedy could have been avoided. Ideally, during a surgical operation, all relevant medical devices should be networked together with reliable safety interlocks automatically enabled. Contextual information in medical records, such as allergies to certain medications, can be routed automatically to relevant devices, e.g., infusion pumps, to support personalized care and safety management. How the patient's vital signs are reacting to medications and surgical procedures can be correlated to streams of imagery data, selected and displayed differently in real time to medical personnel according to their needs, e.g., surgeons, nurses, anesthetists and anesthesiologists. For some particularly difficult stage in an unusual operation, an expert surgeon could remotely carryout the key steps using remote displays and robot assisted surgical machines, eliminating the need for her/him to fly across the country to perform, say, 15 minutes' work. Furthermore, data recording can be integrated with storage management so that surgeons can review the operations and key findings for longitudinal studies for the efficacy of drugs and operational procedures.

2.2 Defense Systems

As reported by Defense Information Daily, "As video communications is integrated into robots, soldiers, and UAVs, and network-centric warfare becomes the organizing principle of American war fighting, front-line demands for bandwidth are rising sharply. The Transformation Communications Satellite (TSAT) System is part of a larger effort by the US military to address this need."[1] TSAT is a part of DOD's effort to build the Global Information Grid (GIG), which is intended to integrate virtually all of DOD's information systems, services, applications, and data into one seamless, reliable, and secure network.

According to a GAO report[2], DOD expects the GIG to enable more timely execution of military operations, collaborative mission planning and execution, common views of the battlespace, and more timely assessments of the condition of equipment and the levels of supplies. For example, greater information sharing could dramatically increase capabilities to rapidly identify and strike time-critical targets, such as mobile surface-to-air missile sites. In the past, such targets have proved to be elusive because the enemy is able to move them to safety in a shorter time frame than it takes U.S. military forces to detect, assess, and attack the targets. By having greater command of a battle situation, DOD expects that lethality and survivability of equipment and personnel would be increased. Armor protection could be scaled down in favor of more agility. In addition, the GIG would reduce the substantial resources and logistics needed to bring command, control, and communications systems to the war-fighting environment.

However, DOD is at risk of not delivering required capabilities within budgeted resources. This, in turn, may affect schedules and funding for other systems depending on the GIG. For example, two key GIG related programs—JTRS and TSAT—are facing schedule and performance risks, which are largely rooted in attempts to move these programs into product development without sufficient knowledge that their technologies can work as intended. In summary, the lack of effective technology for real time embedded network that seamlessly integrates wired and wireless real time network securely and reliably has seriously impeded the development of next generation defense systems.

---

[1] http://www.defenseindustrydaily.com/2005/07/special-report-the-usas-transformational-communications-satellite-system-tsat/index.php#conclusion

[2] http://www.gao.gov/new.items/d04858.pdf

2.3 Tele-presence

As we enter the 21st century, the technological landscape is changing. Looking ahead, we face many challenges. For example, we are now facing the rapid depletion of natural resources, especially oil. In addition, our population is rapidly aging and few senior citizens will be able to afford nursing homes. And the very world we are living in is deteriorating: global warming and pollution threatens the future of mankind.

These seemingly diverse challenges in fact have a common theme. As first articulated by Adam Smith, economic growth is rooted in the increasing specialization and collaboration in production, service and the creation of knowledge. As technologies become increasingly more complex, so is the web of the inter-dependency between specialized productions, services, and knowledge creation. Urbanization was the industrial age's solution for the need for close collaboration. But there is a limitation of how many people can crowd into small spaces. Traffic congestions, pollutions and wasteful use of energy and time due to the need to transport people to and from work have already taxing lives of many in mega-cities.

We need a better solution to meet the grand challenge in economic growth: how to support the ever increasing need of specialization-and-collaboration, without crowing increasingly many people into tiny spaces. The solution is ubiquitous real time tele-operation and tele-interaction that allow human being to transcend the space. The benefit will be history making:

> **Production and service:** Ubiquitous real time tele-operations and tele-interactions would eliminate much of the work-related commute, the dread traffic jams in rush hours that waste energy, pollute the air and waste our time. It also opens the door to allow relatively healthy seniors staying home with remotely performed living assistance. This improves the quality of their lives and will save great sums of money as compared with sending them to nursing homes. It also allows expert surgeons and physicians to provide their services remotely and effectively around the country.
>
> **Research and development**: Currently, productive collaboration in research and development still depends heavily on face to face interactions. The scope of face to face interaction is self-limiting. Currently, the quality of tele-conferencing is indisputably inferior to the quality of face to face meeting that allows not only discussion but also watching body languages and doing experiments together. Ubiquitous and high fidelity tele-interaction and tele-operation will greatly narrow the difference and allow for unprecedented scope in collaboration in research, development and education.
>
> **Public safety**: Local public safety officials face flat or shrinking operations budgets, which limit their ability to add personnel and resources. More than ever before, they are looking for ways to work "smarter, faster, and safer." So are border patrols, and emergency management agencies such as FEMA. A ubiquitous and high fidelity real time space information systems supported by sensor networks would allow effective monitoring and precision dispatching of policy and emergency personnel. Tele-operations will help reduce the number of people and help them to stay out of harm's way.

Many more applications can be found in http://www.telepresence.org/ From the perspective of building the new RTE GENI infrastructure, the key is the creation an integrated backbone and edge networks including of mobile networks with voice, image, sensor and actuation interfaces, which are supported by QoS managed middleware and operating systems. How to create not just the networks, but also the upper level system infrastructures that will allows to securely, reliably, safely carryout tele-operations and tele-interactions is a grand challenge. It is important to remember that the critical challenge to current Internet is not a shortage of bandwidth. The well-recognized security and authentication problem in Internet is QoS problem, not a bandwidth problem. So are the concerns of electric power industry on reliability and security. Remote surgery across hundreds of miles is possible today but the safety concerns stop the idea cold. We need to invent the science and technology to allow for the effective overlay of virtual infrastructures with different QoS requirements onto the same physical infrastructure.

3.0 Summary and Recommendation

To facilitate the collaboration between application community and networking community, it is important to develop a framework to support the collaboration. This framework is called "Requirements-Services Interface (RSI)", which shall include:

- Abstraction of the network properties in the form of a standardized set of QoS service classes that can be economically realized by the underlying networking infrastructure, along with the quality and pricing information of such services.
- Quantization of the diverse application level QoS requirements in such a way that the application requirements can be met cost effectively by selecting from the available QoS service classes.

The main recommendations that came out of the two workshops for the NSF is to sponsor a national level, open, shared experimental facility for the validation of new research results in RTE GENI. As there are several different aspects of real-time and embedded systems, the facility should include:

- Integrated and configurable backbone and edge network infrastructure consists of both wired and wireless network;
- Facilities suitable for RTE application experiments, so that communities including medical, avionics, automotive, process control, defense and security, emergency response, environmental monitoring, power generation and distribution, etc., can experiment with RTE GENI technologies.
- Sponsor the research challenges identified by RTE GENI workshop report.

For the research community, research challenges include:

- Creating networking level primitives that can
    - Provide uniform representation of time and physical location across all networking media,
    - Provide predictable and bounded delay and jitters for communications with specified conditions under which such bounds hold,
    - Support auditing and evidence based certification for safety critical and high consequence applications,
    - Support the evolution of underlying technologies and application requirements,
    - Support automatic fault injection tests, fault identification and isolation mechanisms;
    - Support automatic tracing, diagnosing and reporting of operating conditions and anomalies.

- Creating middleware that
    - Supports key common requirements identified in this report, in addition to the standard middleware services such as naming,
    - Accepts RSI requests from applications, and provides brokering for appropriate services from the network,
    - Support time and event triggers,
    - Provide views of consistent states in real time (resolution may be technology dependent),
    - Provide a uniform interface to access the same type of controls regardless of the underlying network technology,
    - Supports auditing and evidence based certification for safety critical and high consequence applications,
    - Supports the evolution of underlying technologies and application requirements.

- Creating a set of highly reusable protocol modules that can be customized for different application domains. These protocols must
    - Be designed, prototyped and evaluated to reduce the development cost, for specific application domains,
    - Enable interoperability between diverse RTE applications.

Such protocols must handle highly diverse QoS requirements including delay bounds, data rates, geographical distributions, scales, flow characteristics, and safety, security and reliability requirements. For user community with their own networking standards, it is important to continue the participation and collaboration in RT and Embedded GENI efforts and

- Articulate the unique challenges and requirements in specific application areas,
- Participate in the development of RSI and other areas to ensure meeting domain specific needs,
- Participate in the experimental evaluation of the new technologies for their domains to ensure that the common technology can replace one of a kind standard.

References

[1] Profibus http://www.profibus.com/

[2] FieldBus http://www.fieldbus.org/

[3] Bacnet http://www.bacnet.org/

[4] KNX http://www.konnex.org/standard/index.php

[5] LonWorks http://www.fieldserver.com/about/lonworksprotocol.asp

[6] CAN Bus http://www.mjschofield.com/

[7] FlexRay http://zone.ni.com/devzone/conceptd.nsf/webmain/0D17AEEAED870FE486256F3C00407B73

[8] TTP http://www.vmars.tuwien.ac.at/projects/ttp/ttpmain.html

[9] ARINC http://www.arinc.com/aeec/standards/index.html