Smarter and Connected Communities: The REALity Living Labs

Glenn Ricart, Rick McGeer, Andy Bavier, Joe Kochan, Joe Mambretti
August 24, 2015

Abstract

We propose to update the GENI infrastructure to support multidisciplinary living lab research on smarter and connected communities. Smarter and connected communities will depend on response-engineered cyberphysical infrastructure - cyber infrastructure which operates in real-time with the real world. Response-engineered infrastructure is distributed infrastructure engineered to meet the latency, response, reliability, and predictability requirements of a smart community's Internet of Things and its Industrial Internet. We envision cities and rural communities working together with their academic institutions in living laboratories to research, develop, and demonstrate smarter and connected community solutions. The Living Lab projects (infrastructure+researchers+communities) we call REALity. The cyberinfrastructure portion of REALity we call REAL - The Response Engineered Application Laboratory (REAL). It begins with GENI's slicing and deep programmability and reproducibility, and it adds a new emphasis on response engineering including low-latency, reliability, security, and resilience. For the resilience and latency purposes, REAL nodes are located in the academic (or civic) organizations of each participating smart community and are interconnected with the commercial ISPs and carriers of that community. GENI makes prototype REAL infrastructure immediately available in about 50 communities nationwide. The REAL infrastructure should become an NSF MRI intended to support multi-directorate and multi-agency smarter and connected city efforts. GENI's current role as an at-scale computer science Internet testbed can continue but emphasis on response engineering and usability by domain researchers will be added. Over time, a cloud operating system such as XOS should incrementally displace the GENI underlayment.

Introduction: Building on GENI

GENI has all the right initial qualities for REAL: dynamic slicing, deep programmability, and infrastructure virtualization. GENI racks are also located in more than 60 sites across the nation and interlinked with high-bandwidth layer 2 connectivity. US Ignite is interconnecting GENI racks in key metropolitan and rural areas with local gigabit wireless or fiber infrastructure, helping to set the stage for REALity.

Since the core of the REALity Living Labs are the cities or rural communities being served, it is appropriate that the GENI aggregate manager manages racks on a community basis. The local racks also provide considerable resilience for the local community because rack services are not dependent upon remote clouds for their continued operation.
The GENI authentication structure built in InCommon is nicely scalable and amenable to direct use by academics and nonprofits in the community.

While part of a distributed set of GENI nodes, each node is subject to GENI emergency stop procedures to protect other GENI racks from runaway processes and applications. GENI already contains logging capabilities so that usage can be tracked. GENI slices can acquire and dispose of slivers of resources as required to support dynamic cyberphysical systems. In addition, GENI can stitch layer 2 connections between processes and racks to interconnect smart cities and communities.

For big data analysis and computation, GENI can already federate with CloudLab at a deep level, and there is a high priority within GENI/CloudLab to also federate with Chameleon. Assuming permission is granted to allow Chameleon and CloudLab to provide remote cloud services to research and development on REAL, there can be both a local GENI rack with very low latency and a more remote NSF FutureCloud to support highly intense computation and petabyte storage. This hierarchy will enable a wide range of applications to operate effectively in REALity.

Since GENI is often used for at-scale network modeling, GENI has reproducible performance--a key underlay for engineering responsive low-latency and apparently real-time systems.

We plan to use GENI as the initial foundation for REAL. We also plan to assess the steps that will need to be taken to engineer low and predictable response times, inlaid security, and higher reliability into GENI, creating the REAL infrastructure.

Smarter and Connected Communities and the Internet of Things

An area of significant national importance for the research community and for the nation is the emergence of the Internet of Things, and in particular their use to empower Smart Cities and Communities. In practice the Internet of Things means networks of sensors controlling actuators on a scale from the individual device to the city. Some of these sensors are quite high-bandwidth (consider traffic cameras or the array of weather radars in the Cooperative Atmospheric Sensing Apparatus program). Further, real-time or near-real-time reaction to events is a requirement for many Smart Cities/IoT applications, which again imply tight limits on latency between sensor and computing agent and computing agent and actuator. In other words, both bandwidth and latency considerations mandate computing agents close to the action being sensed.

Bandwidth is also important. It is comparatively inexpensive to provide high data rates within a city; the big expense today is taking that data on intercity trips to the cloud and back. In initial pricing experiments, one gigabit city was planning to spend 80% of the gigabit connection fee on upstream fees. What if most of that traffic were kept local? Optical Network Terminals (ONTs) today are inexpensive even at the 10Gbps level and the fiber cost is identical to 1 Gbps or 10 Mbps or 100 Mbps.

Both the ability for local bandwidth and the ability for low latency dictate local cloud capability for applications (and people) valuing responsiveness on the millisecond level. To satisfy these conflicting constraints – tight response from a networked Cloud – we need a Cloud with an architecture sensitive to responsiveness, reliability, and security. We call this a Response Engineered Cloud and it's part of REAL.

REAL, the architecture for research for the next decade, is not just a bigger distributed infrastructure. It is a Cloud Applications Research Infrastructure as fast and facile as the world around us, which can keep up with the demands of Smarter Cities and offer rich Cloud applications with desktop speed. We have already shown an exemplar of such an application: the Ignite Distributed Collaborative Scientific System. The Ignite Visualization System provides seamless interaction and immediate updates even under heavy load and when users are widely separated: the design goal was to fetch a data set consisting of 30,000 points from a server and render it within 150 milliseconds, for a user anywhere in the world, and reflect changes made by a user in one location to all other users within a bound provided by network latency. The system was demonstrated successfully on a wide variety of clients, including laptop, tablet, and smartphone.

This example application is the first representative of a new, class of application, enabled for the first time by REAL. The expectations of REAL are:

- An infrastructure that digests and interprets vast amounts of both historical and real-time data and which is designed to provide apparently instantaneous reactions with the accuracy of being big-data-driven, and renders applications on any device from workstation to tablet.

- Applications that can anticipate human and "smart thing" needs and predict the consequences in both the real world and the --sometimes-indistinguishable --cyber world. Imagine, if you will, a SmartCity emulator which ingests information from a broad array of SmartCity sensors and plugs these values into a continuously running emulation of the city, updating scenarios in the near future and offering warnings and guidance for possible near-future scenarios. The Cloud that manages this application must combine low network latencies, high bandwidth to the sensors, and significant computational and storage power.

- Applications that can help manage and advise us as we humans navigate a complex information-cyber-real-world. The defining characteristic of the applications of the 2010's and 2020's will be that the distinction between these worlds is blurred, and in that merging of the worlds lie the critical research challenges facing our community.

This REAL infrastructure designed to support applications research is the same infrastructure needed for Research and Development on the industrial internet; it is the same infrastructure needed for R&D on Smart Cities, urban sciences and cyberphysical systems. It is an applications research infrastructure that adopts the scalability of PlanetLab, the sliceability and deep programmability of GENI, and adds low and deterministic response times so that it can run in sync with the real world. This is what we need for managing intersections with autonomous vehicles whizzing past at full speed. This is what we need to offer real-time visualizations on a worldwide basis of high-bandwidth sensor data, and to integrate on-the-ground sensors with offline and historic data. This is what we need for in-situ discovery from advanced scientific instruments. This is what we need for quick response to national emergencies. This is what we need for rich, immersive online education such as the Mars Learning Game. REAL is the cyberinfrastructure for the 2020's, and the mission of the research community is to explore it and its implications today. When REAL infrastructure is instantiated on a community-by-community basis to create Living Laboratories, we call that REALity.

REAL as a Cross-Directorate Cross-Agency MRI

REAL will facilitate research that needs not only flops and petabytes, but slices, reliability, security, and predictable responsiveness. The multi-domain needs of smart cities forms an excellent framework to explore the intersection of computer science and interdisciplinary research enabled by REAL.

Directorates	
GEO	Smarter city air and hydrology cyberphysical systems in smarter communities
EHR	Citizen and workforce education in smarter communities
SBE	Social connections and information flow in smarter communities; studies of information economies in smarter communities
ENG	Smart community transportation systems and dynamic systems modeling for most other smart community systems
CISE	REAL infrastructure to support Internet of Things and Industrial Internet
MPS	Mathematical models of smarter communities for prediction, prevention, and remedial actions
BIO	Understanding the balance of nature and man in smarter communities and devising cyberphysical systems to help promote biological diversity
NSB	Coordination of the cross-directorate Smarter and Connected Communities program
Agencies	
DoE	Smart community energy efficiency
DHS and NIJ	Smart community security
NIH and CDC and DHHS	Prevention and early intervention in smart community health concerns
NIST	Interoperability standards for smarter communities and sister smart community programs globally
DoD	Resilience and defensibility of smarter communities

DoC	Economic vitality in smarter communities and building the information economy
DoT	Transportation efficiency and autonomous vehicles in smarter communities
FCC	FCC Model Cities = smarter communities
Dept of Education	Experiential learning via immersion in virtual learning worlds in smarter communities
Federal Consumer Protection Agency	Slices to protect financial information flow in smarter communities
Interior	Bringing the benefits of smarter communities to first nations

REAL thus will become the first NSF initiative since the Internet itself to serve as an MRI across most of the directorates of NSF. REAL will be the cross-directorate MRI for smarter communities research.

In addition, REAL becomes an organizing principle for the National Coordination Office / NITRD since many federal agencies have a stake in smarter communities.

It is proposed that the National Science Board approve and monitor the REALity program as a multi-directorate Major Research Infrastructure (MRI) program housed in CISE, and that NCO/NITRD act as the multi-agency coordination point reporting to the Office of Science and Technology Policy.

Computer Science Research in REAL

REAL can and will continue as a Computer Science testbed for at-scale cutting-edge research, just as GENI has been doing. However, it now has new and added requirements; respond to the real world and real people in real-time. In itself, this is not a new idea. Embedded systems and some real-time cyberphysical systems have been doing this for a long time. But few of these systems have ever operated at the scale of a city or community, and when they have, it's been on a very narrow basis. We've not yet developed the computer science and engineering to integrate real-time systems at the level of cloud computing. REAL will enable such research. And the issues of research and design for secure and reliable infrastructure must be considered at the same time. REALity will help guide the computer science research with realistic problem sets and help promote integration between systems.

A sampling of the interesting research problems to be solved include: (a) hypervising real-time virtual infrastructure; (b) network protocols with deterministic performance; (c) network scheduling for deterministic performance; (d) a robust set of networking protocols with SDN-based congestion avoidance planning instead of packet-dropping (and delay-inducing) congestion inference; (e) ensuring end-to-end security for applications using composed distributed infrastructure; (f) real-time security measures (e.g., no time available for certificate revocation checking); (g) low-power but still real-time secure communications; etc. A workshop (or series of workshops) should be held on computer science questions raised by proposed REALity applications.

Transitioning to REAL

We propose a three-part transition plan to build REAL from the initial GENI base. In many ways, this is like the puzzle where you start off with GENI and change one letter at a time, each time forming a new word, until it says REAL.

```
G E N I
_ _ _ _
_ _ _ _
_ _ _ _
_ _ _ _

R E A L
```

(One solution is given at the end of this paper. Perhaps you'd like to try the puzzle yourself before proceeding.)

Part 1: Community driven technology. REAL Central (See below) should adopt the GENI infrastructure and transition it to REAL under the partnership of the GENI Project Office (through 2016) and US Ignite and the Computing Research Association, with NSF/CISE and NCO/NITRD as the government change agents.

We envision that REAL can quickly leverage current efforts in the networking and cloud research communities. For example, the REAL infrastructure could benefit in the short run from new cloud operating systems such as Larry Peterson's XOS. We need workshops to look at the technical transitions needed.

The Internet of Things will play an important role. Industry and open standards and capabilities need to be integrated into the REAL infrastructure. The Industrial Internet Consortium is a good starting place for appropriate goals and technical standards.

Part 2: Problem-driven approach. Ad hoc groups from the communities in the benefits table will work to propose and adopt grand challenge problems. The grand challenges will be developed in a series of regional workshops in the first half of 2016, modeled on the NSF regional Big Data Charrettes. Since an integral part of this plan is adoption by the GENI campuses and US Ignite cities, CIO representatives from the campuses and the communities, as well as leading researchers from the grand challenge areas, will be recruited to lead the drafting of the grand challenge areas. The National Conference of Mayors will be invited to participate.

Part 3: Birthing Living Labs. City-academic teams will need to work together to conceive and birth the REALity Living Labs in multiple cities. Each REALity Living Lab will develop its own set of challenges and approaches.

Potential Research and Demonstration Applications of REAL

Here are some areas of possible application for REAL and the REALity Living Labs:

1. Smarter Community Resilience. Distributed Applications are robust against local failures and provide failover, particularly for disaster relief. This is an area of extreme interest to NSF and OSTP.

2. As a smart city infrastructure for academic/city collaboration. This is an area of interest to NSF and OSTP: the big idea is that local universities and academics should collaborate around research areas. This implies a Cloud, so that per-project virtual infrastructures can be rapidly spun up. The advantage of GENI is that the management and software infrastructure of this network of local clouds is externally-maintained, relieving the administration burden on local city and campus CIO's. This is one of the major advantages a commercial Cloud provider offers, but it does so at the expense of bandwidth charges and limitations and unnecessarily high latency. GENI offers the advantage of offloaded administration while reducing latency and bandwidth costs and burden. Telemedicine is an excellent example of this.

3. As a way to offload specialized ACI infrastructure. Currently, ACI maintains a network of specialized supercomputing centers tailored for advanced scientific simulations; however, these are often used for more pedestrian tasks, simply because they are the available IT infrastructure. For example, the supercomputer at LBL (need to look up the name) was used for a large map/reduce computation on the CMIP project, after a three-month data marshalling effort over ESNET and other high-speed national and international networks. this could have been done as efficiently on a far-less specialized resource, or collection of distributed resources.

4. As an alternative to data movement. Much scientific computation combines a large data set search with an analysis job on the results of the search. In the CMIP experiment run by LBL and ESNET, astronomers hunt through databases of star and galaxy images for particular phenomena (e.g., standard candles such as Supernovae or Cepheid Variables, or galaxies of specific classifications); physicists for collision events with particular characteristics; geneticists through genome databases. The ability to conduct these searches where the data is collected or lives would be of significant benefit to all of these communities, and is a complement to (3).

5. As a platform for distributed collaboration. This was demonstrated by the US Ignite Distributed Collaborative Visualization System at the Future Internet

Summit in March 2015. It was demonstrated there that this required a GENI-like infrastructure (in other words, a server within 20 ms of any participant), in order for the participant to be able to have a true interactive experience. This was demonstrated as a Pollution Visualizer, but we believe that virtually every scientific community could use such a tool, as demonstrated by the broad usage of the OptIPortal and OptIPuter. Indeed, the Visualizer demonstrates the possibility of "a handheld OptIPortal", at 2-3 orders of magnitude cost reduction over the OptIPortal. As evidence, the Ignite Visualizer is being explored for use in genomics at the University of Victoria.

6. As a platform for the creation of wide-area project-specific Virtual Intranets with tight admission control guaranteed by slicing. This was proposed to the US Government as a key use case for SDN and a demonstration of how slicing could help implement a much more extensive Firstnet at much lower cost.

7. As a platform and framework for international collaboration, both in the domain and computer sciences. We have already begun integration of SAVI in Canada with GENI, and Fed4FIRE in Europe with GENI. We can extend this to Japan's VNode project, and others as the AM API becomes more widespread.

8. As a platform for the continuous modeling and monitoring of Smart Cities and Internet Of Things deployment. Smart Cities and IoT have been described as a "hacker's paradise", since malware and cyberattacks -- and simple bugs -- can have significant consequences. Both pre-deployment emulation and continuous during-operation simulation aided by continual feedback from monitoring of the deployment will have significant impact.

9. As a rapid, secure, resilient distribution platform for IoT software updates. Every company that has to do massive software distribution and updates does this over a CDN, generally a purpose-built one. Steam, a video game cloud distributor, has 243 sites worldwide.

10. As a growable infrastructure for second-life ACI infrastructures which can deliver predictable response times. See example of FOCUS in New Mexico (more details from Rob or Brian)

11. As an efficiency platform/shared cloud for operational cyberinfrastructure across campuses -- Jim Bottum's "condo of condos" (http://condo-of-condos.org/).

12. As a new platform for scientific publication of reproducible results -- not simply the results, but the VM/container with the operating instructions that anyone can use to play with the data, modify or experiment with it. This has been anticipated by Jay Lepreau's last NSF proposal, LabWiki at U Mass, and APTLab at Utah.

13. As a platform for real-time online education. Note particularly the Mars Game demonstrated at the Future Internet Summit. Note, again, that an interactive application on a small device requires a Server Near You.

14. As a platform for more general network and distributed systems research as applied to smarter communities.

REAL Organization: REAL Central

PlanetLab is an excellent model for REAL Operations over the next 3-5 years: the scales and requirements are quite similar. PlanetLab is currently about 1300 nodes at 600 sites; REAL, we anticipate, will have about as many sites and 3x-5x PlanetLab's total node count. PlanetLab was able to both maintain the PlanetLab NOC and develop the PlanetLab code base with a team of 2-3 full time staff and a handful of graduate students, post-docs, and academic researchers, all of whom were doing other things as well. This gives us great confidence that the REAL Operations Center can be managed by an academic group or a small team working for a consortium or nonprofit. We believe that standing up the OC is a high priority. However, we stress that the OC should be modeled on PlanetLab Central rather than, say, the GENI Meta-Operations Center, and should be attached to a research group rather than a collection of operators. Specifically, the OC operators should understand instinctively the needs of the community, because the OC Operators are part of the community. This does not preclude (for example) a campus IT organization; it merely requires that the OC organization have tight ties to the research community, with day-to-day conversation and close understanding of mutual objectives.

In our view, over the long term the growth of REALity will mirror the Internet -- each site will be responsible for maintaining its own equipment and meeting globally-agreed standards. We envision the eventual central organization as being small, and fundamentally devoted to disseminating standards and hosting conferences and meetings, much as the Internet Society is today. This has been the well-established method for infrastructures to grow organically throughout the history of this industry, and we see no reason why this should be different in this infrastructure. In the transition, we believe that the REAL Central Office will not only perform its long-run duties but may also perform the functions of equipment purchaser, and will perform the functions of network operator.

Order of Magnitude Costs

As an order-of-magnitude estimate, we anticipate that the standing costs of operating the REAL/GENI infrastructure will be a centralized cost of 2-4 FTEs for monitoring, operations (such as patch distributions) and front-line support, and an administrative staff of four FTEs: a Director, a Program Manager responsible for purchasing and contracting, an outreach coordinator, and one administrative support person. The figure of 2-4 FTEs is derived from the experience of maintaining Planetlab for a decade. While the PlanetLab Central staff was much larger than that, PlanetLab Central also maintained and developed the PlanetLab code base. We anticipate that the REAL Code Base(s) will be developed and maintained under contract, by groups such as the Flux Lab at the University of Utah, OpenCloud, and ExoGENI, and the REALCentral monitors will function as just that – a small operations center focussed on discovering and reporting problems early.

We anticipate that the on-site labor costs of running a REAL site will be, all in, 20% of an FTE, or 10 FTE-weeks/annum. This assumes three one-week conferences for training, standardization, and clearinghouse activities, and roughly 6 hours/week in the remaining part of the year to respond to node-reboot requests, network and hardware troubleshooting, and installation of software updates.

Capital costs, and in particular minimizing capital expenditure, is one reason potentially to use a central purchasing organization. Volume discounts are available, which would reduce the per-rack cost by up to 1/3, depending upon the volume being purchased. But this would require a single purchaser, who would distribute the racks to the various sites. For InstaGENI, StarLight/Northwestern played this role; for PlanetLab, Princeton/PlanetLab central did. In fact this is a well-used model. These organizations or others could be used as a purchasing agent for aggregating purchases under REAL.

We anticipate a total hardware budget, accounting for one full refresh cycle over the course of the first three years, of roughly $2.8 million or about $1 million/annum. Installation and operations costs would be picked up by the participants.

The networking would be provided where possible by the state and regional networks, or where more appropriate, Internet2. These costs are difficult to estimate, as they are dependent on existing network connections, availability of RENs, etc. However, a worst-case assumption of a new Internet2 Layer-2 port at each site yields a figure of under $3 million/annum

In sum, the order of magnitude for fixed costs of this future are approximately $2.0 million/annum for the REAL Central Office; labor cost over all sites of roughly $3 million/annum; and hardware costs, all sites, of $1 million/annum, for a total of $6.0 million/annum. Networking costs take this to *at most* $8.3 million/annum.

All of the above napkin figures are to be taken as order of magnitude only, given to give the *total* costs over *all* participants, internally- and externally-borne, including already-existing fixed costs (such as existing network connections). In practice, we expect that the total budgeted cost to NSF will be the operation of REAL Central and some portion of the hardware refresh costs, at or under $3M/annum.

REALity Smarter Community Living Labs

The Living Labs will be creatures of the academic-civic-philanthropic partnerships in each smarter community. Leadership should come from the Mayor and university President who appoint a Steering Committee which mobilizes and oversees resources in each REALity Living Lab. Perhaps the National Science Foundation will have a competitive program to help fund some of the projects of the REALity Living Labs similar to what has been done for the Track 1 US Ignite awards. We also expect both monetary and in-kind local support for the changes implemented in the Living Labs.

Transition to Practice

As various projects reach maturity and have independent value to the community, we expect the community will begin to support them. This is the path followed by the original Internet. Also, commercial providers will begin to provide response-engineered hardware and software and the novelty of the research area, after a good number of years, will wane, and new challenges will begin to occupy researchers. This is a normal course of events.

Broader Impacts

The REALity approach offers interesting problems for most areas of Computer Science, and puts Computer Science at the heart of a vast number of interdisciplinary projects spanning much of the reach of NSF and other agencies. Moreover, it's an area of enormous social and political importance. Calit2 Director Larry Smarr is very outspoken on the dangers of global warming, and his take on a solution is very interesting: "Only software can save us": only the efficiencies and responsiveness of the Internet of Things and Smart Campuses and Smart Cities can yield the GHG reductions required without great damage to human society.

(Solution to the GENI to REAL puzzle (don't continue unless you're not going to try it yourself): GENI, GENT, RENT, REND, READ, REAL)