सत्यमेव जयते
Department of Electronics and IT
Ministry of Communications & Information Technology
Government of India

SC2: US-India Smart and Connected Communities Workshop

June 9-11, 2016

(Venue: Indian Institute of Technology – Delhi, India)

Jointly Sponsored by the

National Science Foundation (NSF), United States, and

Department of Electronics and Information Technology (DeitY), India

Organizers:

Sajal K. Das, Missouri University of Science and Technology, Rolla
Uday Desai, Indian Institute of Technology – Hyderabad
Chenyang Lu, Washington University in St. Louis
Huzur Saran, Indian Institute of Technology – Delhi

Workshop Report – June 14, 2016

Executive Summary

The *US-India Smart and Connected Communities* (SC2) *Workshop*, held on June 9-11, 2016 at the Indian Institute of Technology (IIT), Delhi, provided a high quality forum between the United States and India to explore synergistic research collaboration opportunities on various aspects of smart cities and connected communities, a topic of national importance for both countries. Jointly sponsored by the National Science Foundation (NSF) in the US and the Department of Electronics and Information Technology (DeitY) in India, the SC2 workshop brought together 30 thought leaders and visionary researchers from the academic community including 12 from the US and 18 from India.

During the three days of the workshop, the participants were engaged in brainstorming and exchanging exciting ideas on unique research challenges and collaboration opportunities in developing innovative information and communication technology (ICT) solutions, with solid scientific and engineering foundations, that will enable livable, workable, and sustainable connected communities with a goal to improve quality of life. Focused discussions and thematic breakout sessions addressed significant scientific, technological, societal and economic challenges related to mobility, health, safety and security, energy, sustainability, and extreme events management in the context of smart and connected communities.

Specifically, the workshop identified three mutually beneficial areas of US-India collaboration smart cities and connected communities: (1) novel research in relevant key application domains, such as smart healthcare, smart energy, and disaster response and recovery; (2) underlying ICT technologies crosscutting the application domains; and (3) testbeds and datasets at scale that can be shared between researchers in the US and India. This report summarizes the discussions on these ideas, highlights best practices and experiences for effective US-India collaboration, and some recommendations for sustainable collaboration between these two countries.

1. Introduction and Motivation

Advances in synergistic integration of physical infrastructures with networked information systems, smart and sensing devices, multi-modal data sources, pervasive and social computing, and decision-making, all the while humans in the loop, are improving our society and quality of life in cities as well as in rural (or non-urban) communities. In such smart environments, ICT plays significant roles in collecting, elaborating and distributing information and extracted knowledge to the citizens and, at the same time, involving citizens in the creation of information and content (e.g., crowdsourcing). Major research challenges in SC2 include: ubiquitous sensing, communications and computing at scale, dealing with cyber-physical-human convergence, paradigms for collective awareness, knowledge sharing and community building, smartphone sensing, reliability and trustworthiness in crowdsourcing, managing internet of things (IoT), big data analytics, security and privacy issues, among others.

1.1. SC2 Initiatives in US and India

US: In September 2015, President Obama announced new "Smart Cities" initiative and allocated over $160 million federal research to help communities tackle local challenges and improve city services, such as reducing traffic congestion, fighting crime, fostering economic growth, and so on. In fiscal year 2016 the NSF unveiled "Smart & Connected Communities (S&CC): A Vision for the 21st Century" and allocated over $35 million in smart cities related research grants. The vision of S&CC is the effective integration of data sources, networked computing systems, and sensors with people, decision-making, and physical infrastructure to enable more livable, workable, and sustainable communities – *regardless* of place or scale – broadly and inclusively improving quality of life for all citizens in a *diverse* and *heterogeneous* society. The impacts will be far reaching, with potentially transformative applications in important areas such as health and wellness, energy, transportation, environment, public safety, emergency response, and others.

India: The launch of three mega urban schemes in India, such as Smart Cities Mission, Atal Mission for Rejuvenation and Urban Transformation (AMRUT), and Housing for All in urban areas, is aimed to set in motion the process of urban transformation to enable better living[1]. The missions are new, innovative and focused on pressing needs to improve the quality of life for citizens today, and in the future. Currently, 31% of India's population live in cities, and generate 63% of the nation's economic activity. Urban population is increasing rapidly with almost half of India's population projected to live in cities by 2030. India's continued economic growth will be driven by this process.

The Indian government has allocated Rupees 98,000 crore (US$ 15,329.26 million) to implement 100 smart cities and AMRUT, which is an urban rejuvenation program for 500 towns and cities in the next 5 years. In addition to urban transformation, the smart villages initiative, Saansad Adarsh Gram Yojana (Parliamentarian's Model Village Scheme), aims to ensure holistic development of gram panchayats and villages.

1.2. Benefits of US-India Collaboration

The SC2 workshop initiated synergistic scientific cooperation between researchers in the US and India on Smart and Connected Communities, a national priority for both countries. Government initiatives in both countries in this area can benefit one another. There is a strong shared interest in bringing the benefits of SC2 technologies to rural areas (e.g., smart villages in India and broadband access in US rural areas). There is a shared interest in urban issues as well: creating new smart cities and retrofitting existing cities to become smart is an issue that resonates in both countries. The workshop further explored the significant benefits of US-India collaboration in the following three aspects.

Dealing with Diverse Characteristics: The workshop identified a broad set of complementary yet challenging research issues, testbeds, and datasets critical for studying a wide variety of application scenarios and case studies in sustainable societies with integrative ICT solutions. These characteristics are elaborated below.

India is a developing country and has a very high population density (e.g., 950 per square mile) and a highly rural landscape, whereas the US is a highly modern and developed country characterized by large distances and a significantly lower and highly variable population density (e.g., 85 per square mile). Cities in these two countries exhibit widely varying characteristics due to significant socio-economic and cultural differences that are integral part of emerging smart cities and sustainable communities from the perspectives of ICT, pervasive sensing and social computing. Moreover, city centers in the US are mostly financial and business hubs, and they are designed keeping into account human requirements of the 20th century. On the other hand, city centers in India evolved over thousands of years around cultural and social life, and they often lack modern infrastructure from the perspective of healthcare, energy, and disaster response. Such contrasts offer unique opportunities for studying the emergence and evolution of future smart cities by superposing ICT and smart computing technologies over existing city structures. Thus the possibility of investigating future smart cities in these extremely heterogeneous settings will provide the US and Indian researchers a unique set of challenges and opportunities and fertile ground for identifying fundamental as well as emerging characteristics of smart cities. Similar challenges and opportunities also exist in the rural (or non-urban) communities in the US and India, again with significant contrasts in structures and lifestyles.

In terms of connectivity, in India mobile phones and smartphones constitute the most successful technology for connecting the society and accessing the cyber world. India has the third highest penetration of mobile phones having 349 million unique mobile phone subscribers, and about 79 mobile users per 100 people. In contrast, the US has about 98 mobile phones per 100 people.

---

[1] http://www.smartcitiesindia.com/Default-2016.aspx and http://smartcities.gov.in/

Furthermore, India has a rather limited diffusion of (fixed) broadband access at home, with less than 30% of the population having a broadband Internet access, resulting in a significant digital divide across India. On the other hand, in the US, there is a much greater household penetration and access to the broadband Internet.

The SC2 workshop explored the above salient differences in the context of socio-economic and technological organization of these two countries, analyzed emerging properties from the viewpoint of convergent cyber-physical-human worlds, and articulated the scientific foundation for planning, designing and modeling effective smart and connected communities, in particular, providing smart healthcare, energy, and disaster response services.

Sharing Testbeds and Datasets: An effective US-India collaborative research program requires shared testbeds and datasets for heterogeneous cities and non-urban areas in both countries in terms of size, demography, location, and socio-economic settings. Access to certain types of resources and testbeds that are of strong interest to US researchers is easier in India (e.g., large-scale wireless experiments). Furthermore, such a program can provide access to large population groups where research questions are more easily answered in the Indian context (e.g., studies on diabetes, which is a more prevalent disease in India). The SC2 related challenges in India are similar to those found in many developing regions, but a richer ecosystem of research universities and industry base exists in India, which increases the chance of a successful international research collaboration for US academic institutions.

Commercializing Research and Technology: There is a strong entrepreneurship culture and emphasis on IT in both countries (e.g., Silicon Valley in the USA and Bangalore in India). Projects with a translational component or those with potential for commercialization will benefit from this shared culture in entrepreneurship.

Indeed the SC2 workshop provided a forward-looking bridge between the US and India in the area of smart cities and connected communities. Excellent relationship between these two countries at the highest level will help establish stronger US-India strategic alliance in research, education, technology innovation, entrepreneurship, and workforce development to be globally competitive, thus building a stronger link between the two largest democracies in the world.

1.4. Application Focus Areas

The SC2 workshop focused on the applications of ICT to smart healthcare, smart energy, and disaster response. These application domains are not only significant for US and India in building smart and connected communities, but also pose unique challenges so far as the physical environments, demography, social and cultural differences in these countries are concerned. By jointly investigating them will we be able to leverage a broad set of distinct yet complementary research challenges, application scenarios, testbeds, and datasets critical for studying a wide variety of sustainable societies that integrate ICT solutions.

2. Workshop Goals and Activities

This section gives an overview of goals and activities of the SC2 workshop.

2.1. Workshop Objectives

The main objectives of the SC2 workshop are two-fold: (1) to create an avenue for discussions on research collaborations between US and India on the emergent behavior and complex interactions of citizens with interdependent cyber-physical-human systems, such as smart cities and connected communities focusing on healthcare, energy, disaster response and recovery systems; and (2) to exploit this collective awareness and gathered knowledge to develop innovative paradigms for communications and computing, models, solutions and tools to create safe, secure and sustainable societies with better quality of life.

Specifically, the SC2 workshop explored the following questions regarding research collaboration between US and India in the area of smart and connected communities.

- How to build multidisciplinary teams of computer scientists, engineers, domain experts, sociologists and psychologists from both countries to better understand, characterize and model the complex nature of human behaviors and their interactions with ICT enabled smart cities and connected communities?
- How to develop innovative ICT solutions for collecting and processing multi-modal data and making intelligent decisions by exploiting the physical world interacting with the associated cyber worlds as well as humans to create sustainable, smart and connected communities?
- How to design novel paradigms for mining, retrieval and dissemination of large scale heterogeneous data from interdependent cyber-physical-social systems as applied to healthcare, energy, and disaster response systems?
- How to address and quantify tradeoff issues such as accuracy of system performance vs. complexity, security and privacy vs. utility and usability when dealing with smart cities and connected communities?
- How to build multi-institutional and multi-national, large-scale experimental test beds for simulating and emulating smart cities and connected communities? How to generate sharable datasets for a wide variety of experiments and applications?
- How to design effective policies and models amidst socio-economic-cultural boundaries that incentivize cooperation and adaptive interactions among diverse components (e.g., citizens, smart IoT devices, machines, pervasive networking, etc.) supporting smart city and connected community infrastructures?

2.2. Workshop Activities

The SC2 workshop was organized by a four-member team, two from the US and two from India. The organizers from the US were Sajal K. Das (Missouri University of Science and Technology) and Chenyang Lu (Washington University in St. Louis). The organizers from India were Huzur Saran (IIT Delhi) and Uday Desai (IIT Hyderabad). The workshop participants were invited based on their relevant ICT expertise as applied to the chosen focus areas of health, energy, disaster response and recovery. The participants submitted vision statements in advance to the workshop. The three-day SC2 workshop program comprised the following components.

- **Keynote and Plenary Talks:** The workshop featured keynote and plenary talks from India and US that highlighted the current scenarios in each country with respect to smart and connected communities initiatives and identified the most compelling research needs and opportunities for each country.
- **Invited Talks on Prior US-India and Other International Collaborative Projects:** Researchers of two previous US-India collaboration summarized the project outcomes, lessons learned, experiences gained, and challenges faced. Another talk provided an overview of an on-going collaborative project between India and Japan.
- **Breakout Discussions:** Parallel breakout sessions were organized for each of the focus areas (smart healthcare, smart energy systems, and disaster response and recovery). For each breakout session a pair of US and Indian participants were assigned as the discussion lead and the scribe, one from each country. The goal of the breakout sessions was to identify mutually beneficial areas and collaborations between US and India. At the end of each breakout session, the lead-scribe pair presented a summary of discussions to gather further feedbacks from all other participants.
- **Workshop Report Drafting:** The workshop concluded with writing sessions in which the participants collaborated on producing initial drafts of the workshop report which was further refined and integrated by the workshop organizers.

3. Research Challenges and Directions in Application Focus Areas

As mentioned earlier, the broad objective of the SC2 workshop is to investigate collaboratively the emergent behavior and complex interactions of interdependent cyber-physical-human systems in the context of smart cities and connected communities with three focus areas, such as smart healthcare, smart energy, and disaster response and recovery. This section summarizes the underlying research issues and challenges identified at the SC2 workshop in each of the focus areas. In each of the sections corresponding to the focus areas we will present the motivation and current state of the art in both US and India, missing gaps and relevant research challenges for different communities (e.g., urban vs. rural), unique opportunities that can be leveraged from US-India collaboration to advance the state of the art research in the focus areas. Toward the end of this section we will also summarize the interdependency among the focus areas and beyond in the context of smart cities and connected communities.

3.1. Smart Healthcare

Rapid advances are taking place in the area of smart health both in US and in India. The advances in US include mobile (and home-based) monitoring of health using sensors in the phone, on the body, and in the infrastructure (such as homes). Significant work is taking place in developing computational models for sensor data to infer health states and symptoms (e.g., stress, depression, respiratory health, cardiovascular health), daily behaviors (e.g., diet and physical activity, smoking, conversations), and exposures (e.g., mobility, location, pollution). Wearables and fitness trackers are becoming more prevalent and their capabilities are expanding with addition of various sensors on smartwatches (e.g., heart rate, UV exposure, noise exposure, galvanic skin response, etc.). The Big Data-to-Knowledge (BD2K) program of the National Institutes of Health (NIH) is resulting into health data collection and data analytics tools that can potentially be leveraged and adopted to the Indian settings. Precision Medicine Initiative (PMI) is another important initiative that will provide infrastructure platforms, policies, and tools for large-scale data collection, research, and rollout of intervention. This can also be adopted to the Indian settings.

In India, telemedicine and low-cost point-of-care diagnostics are becoming more prevalent to address the very low patient-to-doctor ratio. The point-of-care tools are targeted at various pathogens including those for malaria and dengue. Several centers are targeting betterment of eye health. Another critical need is water portability assessment via point of care technology. In both India and in US, significant work is taking place in geriatrics and independent living.

In terms of communities, there is rapid urbanization taking place in India that is constraining the urban infrastructure, which is making the smart city initiative urgent. While urbanization is a significant contributor to economic growth and prosperity, it imposes numerous burden on the city infrastructure and consequently on the urban population due to worsening congestion, pollution, energy shortage, and overall quality of life. Even though urbanization is causing rapid increase in urban population, significant population still resides in villages, where the scarcity of health care providers is even more acute.

Emerging Gaps in Smart Healthcare in US-India Context: Rapid urbanization, population growth, acute shortage of healthcare providers, and underdeveloped health infrastructure introduce several critical gaps in providing health and wellness to the masses. First, there is a lack of routine health screenings that leads to lack of self-awareness of health and healthcare needs. As a result, most health care encounters happen when health issues become urgent, which increases the likelihood of bodily harm and increases strain on the healthcare system. This points to a growing need for scalable and affordable preventive healthcare.

Second, the data collection and labeling is prone to errors and other inaccuracies (e.g., provenance) due to underdeveloped skill set of health care workers who outreach to the population to address the acute shortage of trained healthcare providers. Lack of scalable training tools (e.g., simulators) is another bottleneck in rapidly scaling the size of trained healthcare workforce to match the increase in population and their healthcare needs. Smart health technology is needed that can help collect, screen, clean, and process the health related data so the first stage of triaging can be done before involving trained health care workers.

Third, several technologies developed in developed countries are not directly applicable to the Indian population due to differences in genes, behaviors, and the living environment. Hence, cohorts of Indian population is needed that can be used to adapt and test emerging smart health technologies so that they are applicable to Indian population, which represents a large and rapidly growing market for such technologies.

Fourth, given the developing nature of Indian economy, low per capita income, and large population base, smart health technology needs to achieve very low cost and extreme scale to be applicable and adoptable in the Indian settings. Open and modular sensing platforms can help spur innovations to reduce cost and improve the scale of these platforms so they are suitable for the Indian population.

Finally, Indian population consists of both a large group of young and tech-savvy population and undereducated population, both of which need health care. Such a mix of population represents both a challenge and an opportunity to develop smart health technology that can leverage the human network and become accessible to the population as a whole irrespective of their technological sophistication.

Unique Research Challenges: Collaboration between US and India will represent numerous opportunities to advance science as well as improve the health of a large section of world population, as described below.

- Rapid urbanization in India offers an opportunity to study its impact on the health of its residents. Understanding the impact of urban planning on health of its residents represents another interesting and societally useful research problem as it can inform the urban planning that can be optimized not only for transportation and energy efficiency, but also directly targeted at better health of its residents.
- Development of smart cities offers an opportunity to investigate scalable and affordable smart health infrastructure than can be shared by the community for regular health monitoring and preventive health care. A potentially scalable approach may be to develop smart ambulatory healthcare units that can visit patients, perform routine screenings, obtain results of screening in real-time to make care decisions, and if necessary handover to an appropriate healthcare provider on the spot (otherwise, the patient may fail to follow up or may have difficulty reaching out to the appropriate health care provider).
- Development of smart transit care (e.g., ambulance equipped with smart healthcare technology) can help save many lives by caring for patients on the way to hospitals via ambulances. This is important in both the urban and rural context, but especially lifesaving in the rural context, but is challenging given the transportation context and limited capability of the transport medium.
- Providing personalized health assessment from these community owned technologies (e.g., personalized measurement of pollution exposure from pollution sensors deployed throughout a city by fusing GPS traces) represents interesting research challenges. Resolving the identity of the individual whose data is captured by the smart health technology (e.g., sharing of smartwatch or smart toothbrush by multiple members of the same family) represents an interesting research problem.
- Smart homes in the Indian context are usually co-inhabited by multiple patients, often with the assistance of a health care worker. Due to the sharing of living space, there may be conflicting demands on the infrastructure. Developing features in such multi-user smart homes that can provide personalized experience despite the sharing of both the living space and the smart healthcare technology embedded in it represents interesting research challenges.

In addition, to achieve scale in mental health management and elderly care, novel smart health care technology is needed so that one healthcare worker can effectively and efficiently manage multiple patients concurrently, thus achieving scale. For example, providing a means for a healthcare worker to simultaneously interact with (and potentially control) assistive technologies (e.g., brain computer interfaces) being used by the patients they are managing can help achieve economies of scale while providing the appropriate care needed by their patients.

- Modeling of daily behaviors in the Indian context requires significant new research. This includes sensor-based methods for nutritional analysis of dietary intake (e.g., higher prevalence of salt, fat due to deep-frying and widespread use of milk-based food, and sugar in sweets) and tobacco consumption via chewing of gutka. Given wide prevalence of risk for cardiovascular diseases (CVD) in India (i.e., over 70% of Indian population is at the risk of CVD and salt, sugar, milk, oil, and fat rich diets, and tobacco are significant risk factors for CVD), monitoring of these behaviors is critical in providing preventive health care.
- Prevalence and significant expertise in yoga and meditation in India and its increasing popularity in the US offers an opportunity to develop innovative cognitive behavioral therapies (along the lines of mindfulness) that can be delivered remotely (e.g., on mobile devices).

3.2. Disaster Response and Recovery

Disaster response and recovery was identified as a key area of interest for US-India collaboration. In the US, losses due to natural disasters alone totaled $16 billion in 2015, whereas, in India, disaster damage is estimated at about $10 billion annually. Together, India and the US therefore account for more than 30% of the global world-wide disaster cost, estimated at $85 billion in 2015. The motivations for collaboration on disaster response between India and the US are therefore self-evident.

Research Challenges: The workshop emphasized the need to focus on urban (smart city) and rural settings and distinguish at least two disaster types; (i) disasters that impact the physical infrastructure, such as earthquakes and hurricanes, and (ii) disasters that do not, such as infectious disease outbreaks. The two types offer somewhat different challenges, calling for different technology solutions. Together, the following research challenges has been identified:

- ***Medium access challenges:*** Dynamic spectrum access and management of spectrum sharing were identified at key challenges. Of particular interest is management of public safety bands reserved for primary emergency use. Normally, in the absence of disasters, such bands will not carry disaster traffic, making them available to other parties, via appropriate bidding mechanisms. When a disaster situation arises, present band occupants would be denied in favor of emergency traffic. A key issue in this context is to offer a balance between guarantees offered to disaster management traffic and predictability offered to other traffic. Algorithms for cognitive radios are needed to support viable economic and business models for sharing this band.
- ***Network-layer challenges:*** Significant amounts of work can be leveraged on mobile ad hoc networks for emergency response. While a large research investment was made in delay-tolerant networks (DTNs) and mobile ad hoc networks (MANETs) in the past decade, several issues remain open in the context of maximizing information transfer for disaster survivors. When infrastructure is significantly disrupted, novel solutions are needed to break away from the limits of (impaired) channel capacity. They may include new networking paradigms, such as information-centric networks, that enable the network to do more efficient triage, as well as information-value-based congestion control that leverage shared prior knowledge between producers and consumers to increase the rate of useful information transfer. Network management, configuration, and adaptation remain open problems.

- ***Data broadcast:*** When conventional infrastructure is disrupted, unconventional data broadcast channels may be used, such as TV data-casting that exploits TV broadcast media for data communication. Exploiting TV data-casting for disaster response operations offers additional research challenges. For example, how to adapt the data-casting content depending on current needs? How to incorporate feedback into the inherently one-way data-casting framework? How to allow survivors and rescue workers to exploit the framework for two-way information exchange?
- ***Crowdsourcing/Crowd-sensing:*** This is a key capability in disaster management because it does not rely on new infrastructure and so does not need a financial investment. Feature phones (with SMS capabilities) and low-end smartphones (e.g., with a camera) are sufficient for implementing a large array of crowdsourcing applications, thereby empowering the human to act as a "sensor". Leveraging the humans in the loop is a key promising capability to develop in the context of disaster management.
- ***Social network exploitation:*** Exploiting social networks for disaster response has proven to be a successful practice in media-rich environments, such as in Japan and in the US. An interesting question is how to extend the advantages brought about by social network exploitation to environments with a less abundant technology penetration profile? What mechanisms are needed to spread situation awareness during and after disasters? Networks such as Twitter are compatible with SMS, and SMS capable phones have wide technology penetration in India, totaling over 1 billion devices. Can systems be built that offer social networking capabilities customized to local needs on top of widely available lower-end phones? In an environment with many local dialects that make text input cumbersome, what are suitable alternative modalities for social information dissemination?
- ***Real-time and Diagnostic Data analytics:*** When exploiting unreliable information sources, such as unveted participants in crowdsourcing and social media exploitation applications, advances are needed in data cleaning and analysis to offer a reliable automated understanding of key events, as well as their locations and attributes, to support informed decision-making for disaster management. Real-time analytics becomes challenging to speed up comprehensive event analysis in real-time and empower first responders in disaster management.
- ***Human-machine teams:*** Other challenges identified included middleware and applications that allow exploitation of robotics (e.g., using drones to set up communication networks) as well as effective communication between robots and disaster response teams or disaster survivors.
- ***Internet of Things (IoT) opportunities:*** Opportunities of using IoT technologies in smart cities is promising for urban settings. Sensing, communication, and computing capabilities for IoT-based disaster management need to be investigated to enhance efficacy of disaster management in smart cities. The research challenges cross design layers including sensing, analytics and communication.

Indo-US Collaboration: In addressing these challenges, a US-India collaboration would offer several opportunities that could lead to integrated application-aware solutions, where expertise is assembled from the physical layer, to the network layer, systems, middleware, and applications. In assembling teams, two needs were emphasized. First, it is important to incorporate domain experts that have more familiarity with the underlying physical phenomenon of the disaster. Second, it is important to include social scientists who understand the population. This is because a true solution to disaster management must not only address the information technology component, but also consider the physical and social context in which the technology solution operates. Hence, teams are encouraged to include the appropriate experts on physical/environmental, social, and information technology aspects of proposed solutions.

3.3. Smart Energy and Buildings

Smart energy and buildings are key elements of future sustainable connected communities. Many research activities in both US and India have focused on creating technologies to make electricity grids more efficient, reduce energy consumptions, and make buildings smarter. In this workshop, we identify unsolved and new research challenges, with respect to the strengths of each country, and propose new research areas to focus efforts.

Research Challenges: The following research challenges have been identified.

- **Smart Metering Analytics:** Recognizing the importance of real-time electricity use metering, utilities in many urban areas have deployed smart meters over the past decade. While most of the sensing and communication aspects of smart metering have been addressed, many challenges remain in analyzing and taking advantages of the wealth of smart metering data to provide tangible benefits for utilities and customers. Techniques for analyzing such data while ensuring user privacy is also important. At the building level, correlating occupancy data to energy use to provide intelligent subsystem controls is a rich research opportunity.
- **Load Shaping.** Peak shaving and load shaping are two promising ways for demand-side energy management and reduction. While a number of research efforts have focused at the feeder and building scale, relatively few efforts have focused on managing loads at a finer granularity. For urban communities, fine-granularity load shaping in both space and time presents new opportunities for significant energy savings. With greater intelligence embedded in buildings, for example, awareness of occupancy levels and preferences of occupants, buildings can adjust various loads dynamically to shape its consumption profile. At the opposite side of the spectrum, coordinated load shaping across multiple buildings provides additional opportunity for energy reduction, but also new challenges due to heterogeneity of assets and control systems.
- **Interoperability.** Equipment and systems inside electric grids and buildings are typically developed by corporations such as ABB, Siemens, and Schneider Electric using proprietary interfaces and protocols. It is very difficult to interoperate between systems made by different vendors. Research on creating unified interfaces and standard protocols for these classes of systems could facilitate innovation in smart grid and buildings.
- **Actionable Data.** In smart buildings research, significant progress has been made on the sensing (both direct and indirect), networking, and visualization of energy usage data inside buildings. However, while it is informative to have fine-granular visibility into the power consumption of buildings, the amount of actionable data is still very limited. For building managers and occupants, it is often more important to address the question of "what can I do to reduce energy?" This presents interesting research challenges, including additional sensing modality and analyzing data from the occupant's perspective. In addition to systems challenges, economics, behavior, psychological, political, and policy also play significant roles.
- **Security and Privacy:** Given the prevalence of cloud-based IoT devices for smart buildings, the privacy of user data in the cloud as well as the security of remotely accessible home devices and appliances is a critical problem. A unique security problem in India in the loss of electricity due to theft, which prevalent in both urban and rural communities. Research challenges exist in both detection and prevention of electricity theft.

US-India Collaboration: US and India share many common problems and characteristics. For example, both US and India have high urban populations and remote rural areas. Consequently, energy optimization in the face of climate change is a priority for both countries. Technologically, both US and India are IT powerhouses, with strong talent pools for developing smart energy and building technologies. Perhaps more importantly, US and India have many differences in terms of challenges and strengths, and can complement each other in various ways. Because of the scarcity of power, India has a dire need for intelligent load shaping to increase reliability. In the US, with time-of-use (TOU) pricing, peak shaping has proven to be effective in reducing electricity costs. Developing effective load shaping technologies is important for both countries. In the US, there are significant research results on how to interface with, and collect data from, building management systems; in addition, there are several products on the market for smart home automation. These technologies have significant potential for deployment in India. On the energy storage side, battery storage in the form of UPS is very common in India. This provides an excellent testbed for innovative research in load shaping in both India and US. The US has developed various off-grid technologies, such as those utilizing solar and storage, which can be adopted by India where the electric grid may be completely missing in isolated rural areas.

With these differences and complementary strengths, collaborative research efforts should be focused on the following areas:

Human in the Loop: In addition to monitoring energy use, research should focus on developing smart buildings with humans in the loop and using data analytics to produce actionable feedback to occupants, ultimately motivating and incentivizing occupants to conserve energy.

Load Shaping via Energy Harvesting and Storage: With 24x7 zero energy communities as an ultimate goal, research should focus on developing technologies for load shaping via innovative use of renewables and storage, potentially exploiting existing residence based batteries already in India.

Off-grid Energy Management: with increasing number of self-sustainable communities in the US and prevalence of rural areas without grid in India, research and technologies for off-grid energy management, including developing new models for resource and load prediction and energy storage, will have significant impact in both nations.

Micro Spatio-temporal Load Reduction: Energy wastage in large shared spaces such as commercial buildings, malls, and schools can be reduced significantly by using more intelligent sensing, data analysis, and control. Research should focus on micro-scale load reduction at fine granularity of space (meters) and time (seconds or less), potentially leveraging existing indoor positioning technologies in US and India.

3.4. Cross-Cutting Research Challenges and Issues

Smart and connected communities need to integrate multiple capabilities seamlessly, which requires research at the intersection of the focus areas of health care, energy, and disaster response, and how these are shaped by and affect human and societal behavior. Human behavior is central to how people use various services and resources including health-care and energy and is central to successful mechanisms for providing the services and influencing their resource usage. The same holds for providing disaster response, which can be considered as a service. The centrality of human behavior was noted in all three tracks. However, there are behavioral issues that go much beyond these connections. One key aspect is that human behavior itself needs to be examined at multiple levels because of social relationships and people to people interactions. In particular, providing health-care to a community needs to go beyond providing the same to an individual. Similarly, the decisions that a person makes regarding energy usage and how those decisions are influenced by energy-saving or peak shifting incentives is strongly influenced by the decisions that others make in the community. Thus, understanding social dynamics and how it influences the choices, and is driven by the available alternatives is crucial to smart health-care and energy management.

Disasters affect human and societal behavior in complex ways; it is well known that people in a crisis situation (or under the threat of a crisis) behave very differently than under normal circumstances and this is crucial in understanding how services should be provided to them in those situations. In particular, individual and community smart health-care solutions that work well under normal circumstances may be entirely inappropriate in disaster situations.

There are other cross-cutting themes as well that need to be considered. Disasters not only change the human behavior of the people within the disaster zone, but also have direct impacts on the people (injuries, entrapment, etc.), infrastructure (damage/physical inaccessibility to the health care facilities and infrastructure), and logistics (for getting equipment, medicines to people, transferring them to specialized care facilities, etc.). These aspects make the health-care provision in a disaster situation markedly different from that for normal circumstances.

Disasters disrupt not only the communications networks but also energy networks (both electric and others such as gasoline for transportation). The increasing “smartness” of the physical network increases its dependence on the communications, which themselves may be dependent on the energy network. Understanding this coupling and how the disasters play into this is a crucial element to making energy networks resilient. As is well known, the integration of distributed generation resources (e.g., multiple local solar and/or wind farms) into the power grid has its pros and cons (distributed generation => More resilient to failures but integration of variable sources => more instabilities).

Disasters could result not only from natural or technological mishaps but also caused deliberately by intelligent adversaries, particularly through the increasingly stronger cyber route (e.g., bad data injection in smart grid). Understanding the impact of such disruptions on energy networks, and the resulting impact on health-care are other crucial issue for making both energy systems and health care systems more resilient of disasters.

Research Challenges for Urban/Rural Communities: Understanding human behavior at both individual and community level and exploiting it for the resilient design and efficient operation of energy and health care systems is an open problem that requires close cooperation with social scientists.

The telemedicine capabilities provided by ICT can be helpful in disaster scenarios if designed carefully by considering its limitations (as compared to in-person health care). In particular, we need smart positioning of mobile communications and health-care assets in the disaster area (both land based and aerial), suitable sensing capabilities, further backed up by analytics to determine what capabilities can be supported.

The increasing use of networks of micro grids driven by renewable energy need to be designed and used with disasters in mind for a good balance between demand and supply. On the supply side, the grid may be disrupted and stabilizing the partitioned system in the presence of distributed generation resources (e.g., disconnected micro grids) is a challenge. On the demand side, it is necessary to prioritize usage, for example, giving higher priority to health-care and subsistence. Prediction and management of health issues arising due to the effects of disaster (water contamination, food spoilage, lack of prevention capabilities, epidemics) are other open research challenges.

Since disasters may destroy part of sensing infrastructure as well, there is a question of how to best reposition remaining sensors to provide the functionality and what is feasible. A software defined sensing approach is useful here but needs to be thoroughly explored.

US-India Collaboration: The US has much lower population density than India, more resilient infrastructure (less likelihood of mass injuries), and much greater capabilities for evacuation of sick/injured via helicopters. Thus, remote health monitoring technologies are crucial in India, but could be relevant for US as well.

4. Research Challenges and Directions in Underlying Technologies

Research on SC2 is beyond individual application domains. The workshop participants identified three key technologies that underlie and crosscut different application domains: (1) sensor networks and cyber-physical systems, (2) wireless communications, and (3) data analytics. To advance the state of the art in SC2 research both application domains and underlying technologies must be considered. This section identifies research challenges in the underlying technologies.

4.1. Sensor Networks and Cyber-Physical Systems

SC2 are complex human-cyber-physical systems where multiple sense-infer-decide-act feedback loops, nested and distributed at different spatiotemporal scales, work in concert to provide a range of services. While such sense-infer-decide-actuate feedback loops were a recurrent theme during the various application-focused sessions at the workshop, the discussion rapidly converged towards the viewpoint that a SC2 will not be an assembly of a multiplicity of bespoke hardware-software stacks for different services. Instead, it will be built upon the foundation a shared and programmable city-scale substrate with integrated sensing, processing, communication, and actuation capabilities on top of which an ecosystem of “apps” can be instantiated as necessary to serve specific needs of the citizenry. How best to organize and implement such a substrate, particularly one that has necessary scale and robustness, is challenging. On the one hand, the key architecture principles, abstractions, interfaces, and layering are not yet properly understood while the limitations of simply applying extant architectures are evident. On the other hand, many of the building block technologies necessary for the substrate are not yet available.

Much of the discussion at the workshop focused on the sensing functions in the substrate. Most SC2 services depend on knowing, and even anticipating, the dynamic state of a community, a capability which in turn depends on providing actionable information at the right time to the right place in a SC2. The discussion highlighted several challenges, some of which are presented below.

Fusing Diverse Sensory Data Sources: SC2 would have multiplicity of “hard” and “soft” sensory source. The former category includes embedded sensors that instrument physical spaces, as well as sensors embedded in smartphones and wearables in myriad form factors (smart watches, smart earphones, and smart glasses). The latter category consists of human sourced information such as from social media postings and contexts. Such fusion, while essential to achieving low-cost, coverage and resilience, is challenging because of factors such as variable quality, lack of calibration, natural language ambiguities etc. Of particular importance would be development of sensing methods that require minimal instrumentation so as to be burden-free for users and operators while keeping capital and operation costs acceptable even for cost-sensitive deployment scenarios such as villages and rural areas.

Pushing Sensor Data Analytics towards the Edge: Current networked sensing systems dominantly follow the model of sensor devices at the edge collecting data and sending it to cloud-based analysis services for rich context and behavior inferencing, and for triggering actions. While sufficient for small-scale deployments, such a model will not scale to city-scale systems with hundreds of millions of sensors. The sensor data deluge at such scale will overwhelm both the communication infrastructure and the cloud-based processing services. Crucial to scaling in SC2 would be to push processing of sensor data to the edge, which would significantly reduce dimensionality of sensor data and enable smarter data collection based on its information value. Accomplishing this would require innovations such as energy-efficient and lightweight stream processing and machine learning algorithms suitable for embedding in edge devices, and accompanying advances in sensing platform hardware. Eventually we expect the processing of sensory data in SC2 to be flexibly and dynamically distributed along the data-to-decision pathway spanning the edge and the cloud while adapting to available connectivity.

Scalable Sensor Metadata Management: For sensor data to realize its full utility, it must be placed in context via metadata that annotates it. Besides common metadata such as timestamp and location, metadata can also capture parameters relating to the quality of information and calibration. Current approach based on manually annotated metadata is not only error prone but also not scalable to large scale system. Additionally, lack of standard representations and algorithms to meaningfully handle metadata prevent an end-to-end treatment. SC2 will require metadata to be treated as a first-class entity that is processed through stages of sensor data processing as it flows through the system, and allow it to be queried and searched for purposes such as forensics, provenance, and fusion.

Secure and Privacy-aware Sensing: Making sensing trustworthy, both for producers and consumers of sensor data, would be essential in SC2. On the one hand sensors can capture sensitive information, creating privacy and safety problems for those being sensed. On the other hand sensor data can suffer from bias or its integrity intentionally compromised, and in turn influence decisions and actions that are based on using the data. Unlike normal data stored by our communication and computer systems, the physical nature of sensor data expose them to a much broader range of risks of errors and attacks along cyber and physical vectors. Devising sensing methods that are both secure and privacy aware would thus be a key challenge. In SC2 scenarios protecting the privacy, identity and integrity of sensor data streams is made particularly challenging by the multiplicity of stakeholders, different sharing policies and trust relationships, and varied social norms and legal requirements.

Making Sensing Manageable: Another challenge to scalable sensing in SC2 is the management of the substrate consisting of hundreds of millions of sensors. Besides monitoring the health of the sensors, the management tasks also include making sure that battery-operated sensors do not run out of energy and that their software is up-to-date. Meeting the management challenge will require innovations such as high-efficiency energy harvesting technologies and scalable methods for monitoring and updating large-scale network systems.

New Sensing Middleware Services: For apps built on top of the SC2 sensing substrate to effectively use the collected information, middleware functions such as for searching and processing in real-time a large number of high frequency streams of sensory data would be important. By raising the semantic level of sensory data from raw measurements to query responses and rich context and behavior inferences, middleware services would enable efficient use of SC2 sensing substrate.

The sensing functions in the substrate are complemented by communication, processing, and actuation capabilities. The challenge in the first two are discussed in depth in the two ensuing subsections, and therefore here we discuss the actuation capabilities via which the feedback loop is closed so as to deliver interventions that nudge the community and its citizens towards a desired state of health, energy efficiency etc. One common form of intervention is informational where just-in-time alerts, suggestions, feedback etc. are delivered to the end user in an attempt to influence their behaviors relating to health, safety, and sustainability. There are many examples of such systems in various domains, such as air quality warning, earthquake warning, health related reminders etc. but these are one-off isolated systems. The SC2 substrate could provide management, scheduling, and delivery of interventions as a shared systems service. A challenge in doing so arises from human factors relating to differences in language, cultural norms, literacy levels etc. across the population. Beyond behavioral interventions SC2 apps will also make use of physical interventions whereby traffic lights, building systems, and other resources in a SC2 may be controlled based on sensory information. Such interventions would present the challenge of ensuring timeliness and resilience in networks operating under extreme scale and suboptimal connectivity.

4.2. Wireless Communications

Wireless communications research is critical for developing healthcare and disaster management technologies in smart and connected communities. Telemedicine for healthcare and remote monitoring for disaster recovery are two important examples that support this observation.

Cognitive radio networking enabled dynamic spectrum access, cloud offloading, and mobile cloud computing present another primary research opportunity. Research in this area can enable critical technologies and applications such as smart (mobile) cloud computing for resilient communications, long range rural wireless broadband for telemedicine, middle mile/last mile access that bring connectivity as an enabler of smart villages, and mobile healthcare for rural communities. A few research topics concerning these applications are:

Software Defined Networking: programmable radios that can autonomously reconfigure themselves depending on the operating environment or applications' needs; application aware radio and computational resource allocations; network scalability; flexibility to mix and match the underlying technologies easily are a few desirable features. Software defined networking is an enabler of research in this direction.

Resilient and Secure Network Protocols: communication protocols must be resilient to different types of impairments such as power outage, wireless coverage due to mobility, disaster situations, etc. Where majority of the systems in SC2 are assumed to be autonomous it is important to investigate the impact of security vulnerabilities of network protocols on the overall system stability.

Heterogeneous Networking: SC2 will most likely have heterogeneous networks consisting of a mix of WiFi, 4G/5G, fiber, and other types of physical layers. Therefore, managing these intelligently and in real-time, perhaps from a distributed cloud infrastructure, is critical. Decision making and learning algorithms, novel control theory, game theory and other research techniques may be applicable to solve this problem.

Research in 5G mobile communications and the Internet of Things (IoT) within the SC2 context has to address unique needs of emergency response and healthcare operations. This includes enhanced *device to device (D2D)* and *massive machine type communications (MMTC)*.

A model city or village for testing at-scale, next generation wireless communications research for healthcare and disaster response must be developed. Such a unique, shared infrastructure will benefit the research community significantly. It will also benefit companies that could partner with the research teams for translating basic research into commercially viable products.

4.3. Data Analytics

The collection of sensor data form the first stages of a long data flow pipeline towards the production of useful and usable information for sophisticated SC2 applications. In each one of the application areas (healthcare, energy, and disaster management), small scale applications can be built on top (or into) smart sensors. However, truly smart applications need significantly improved information, both in quantity and quality that require new breakthroughs in big data analytics, e.g., novel filtering and integration algorithms and tools. In addition, sensor data also need to be analyzed in its appropriate context potentially defined its spatial, temporal, "socio-economic" coordinates. Ignoring this contextual information could lead to misleading and potentially incorrect results.

To illustrate the challenges in big data analytics in SC2, we first provide a few sample datasets typically generated in a smart city.

Video Streams from CCTVs: As an example of high volume real-time data processing requirements in a smart city, Hyderabad is installing thousands of video cameras for street monitoring. These cameras (up to 8 per street intersection) produce high definition videos (several mbps per source) of street scenes that can provide useful information on traffic incidents and actual crimes committed in public spaces[2]. Processing such massive video data and detecting events of interest automatically (both in real-time and non-real-time) is an expected service of smart cities, and it presents significant challenges for the current generation of data analytics tools and machine learning algorithms.

Data from Social Sensors: In addition to physical sensors, a growing source of information is social media. This is particularly the case in the cities due to high population density, pervasive smartphones, and good networking support. Human reporting of events can become useful sources of information, either primary or supplementary, for large scale events such as natural disasters or

[2] *The Telegraph (Jan 1, 2009). Scotland Yard study: 7 out of 9 murders are solved with the help of CCTV video. [http://www.telegraph.co.uk/news/uknews/law-and-order/4060443/Seven-of-ten-murders-solved-by-CCTV.html]*

human-made emergencies (e.g., the Boston Marathon Bombing). These “social sensors” (e.g., Tweeter, YouTube, Facebook, and Whatsapp) contain significant amounts of information, often unavailable from physical sensors, but in compensation, the social sensors also contain a large amount of noisy data.

Geospatial Datasets: Another example of sensor data of special interest to SC2 applications is geo-spatial data, e.g., datasets containing space and time coordinates that describe physical events. Concretely, location data are often included automatically by smartphones (for social sensors) and many health sensors. Typical sensing frequency of these sensors is of the order of 1Hz, there by easily reaching sizes of terabytes or petabyte for a city scale dataset. Geospatial datasets can answer questions like: “Where are the traffic congestion points/routes during rush hours?” “Can we find this car seen leaving a crime scene 5 minutes ago from street monitoring CCTV cameras?” “Can we create a map of real-time population distribution and evacuation routes for large (and small) disasters?” “Is there a statistically significant correlation between space-time coordinates of where use of asthma inhalers were used and pollution levels?” “Are there any anomalous regions where the energy consumption has suddenly increased (power theft and illegal installation of heavy machinery on domestic power supply) and is significantly higher than their neighbors?” These questions involve static historical analysis (traffic congestion, usage patterns of electric power and asthma inhalers in a city), real-time video data analysis (fleeing car), and combination of several data sources and sensor data (real-time evacuation plans).

Despite value addition potential of smart city datasets, they pose significant challenges to the state of the art in data analytics. Some of the challenges are listed below:

Noisy Data: Noisy data in social sensors can be roughly divided into two categories: (1) unintentional noise, e.g., a “landslide victory” in an election contains the same word as a landslide disaster; and (2) intentional spam, e.g., unfounded “rumor tweets” that appear during every major event, often associated with some kind of promotion. For both unintentional noise and intentional spam, *a priori* cleaning of social media data is often difficult or impossible, particularly within each social media channel itself. This is due to the inherent ambiguities of natural language text (in the unintentional noises), and the purposeful imitation of legitimate content by intentional spam. The filtering of such noise is another example of significant challenges for current data analytics tools.

Balancing False Positives and False Negatives: Classic machine learning algorithms and tools have some significant limitations when filtering the various kinds of noise present in physical and social sensor data. An example of such limitations is the trade-off between false positives and false negatives when trying to decide whether a data item is relevant or irrelevant to the topic of interest. Very often, trying to minimize one kind of error (e.g., false positives) would raise the other kind sharply (in this case the false negatives). Consequently, statistical analyses and machine learning tools are applied on the brightest signals in areas such as trend analysis, where the large number of data items facilitate the retrieval of relevant information. For SC2 applications, trend analysis works for a subset of situations, e.g., when large disasters strike and become trendy in social media. In other situations where the topics of interest are less popular, data analytics techniques tools need to be improved or new techniques developed to find such fainter signals.

Studying Data in Appropriate Context: SC2 question and answer require data analytics techniques which study the data in its appropriate context (space, time, space-time, socio-economic) and pay attention to its appropriate syntax and semantics. For example, classic algorithms often make assumptions such as the independence of sample data sets, and they work primarily on uniform data sets of similar syntax and semantics. This conflicts with the location semantics inherent in a geospatial dataset.

Fusing Data from Multiple Sources: Another challenging research area of data analytics that offer promise (and may be required) in SC2 applications is the integration of data from a variety of sensors, including physical and social sensors. There exist many applications that can benefit from such integration. For example, air pollution from car emissions (data from air quality sensors) may be related to road congestion (data from video and Google Maps). Conversely, there are significant challenges in the integration of such varied sensor data, due to their differences in syntax, semantics, noise levels, and other interactions among the data sets and sources.

In summary, SC2 offer a great opportunity for data analytics research, and they require significant advances in data analytics techniques and tools for SC2 to fully benefit from the variety of sensor data that will become available. As an example of concrete research challenges, significant new capabilities in the integration of varied data sources, from both physical and social sensors, will be required to make SC2 of the future meaningfully smart.

5. Common Testbeds and Datasets

An important benefit of US-India collaboration is to share testbeds and datasets with diverse characteristics. In this section we highlight opportunities to share testbeds and datasets between both countries to support SC2 research. For each application domain, we first identify the need for SC2 testbeds and their desired features and related challenges. We then describe what components exist today and what new developments in testbeds are needed. Finally, we highlight opportunities to share them between US and India.

5.1. Smart Healthcare

Motivation for SC2 Testbeds for Smart Health: In the context of the smart communities (urban or rural), it is crucial for us to understand not only how the community infrastructure and environment impacts health/wellbeing but also how the community can improve health/wellbeing of all its participants. Considering the complexity of what makes a community, e.g., transportation, energy, weather and air quality, the complex inter-relationships between community and health requires data that cannot be faithfully captured by a small-size testbed. Thus, the scale is crucial for truly understanding the relationships between where we live and work, and how it impacts our physical and mental health and wellbeing. With scale, one can ask and answer following forms of new questions:

- Researchers, public health officials, and community planners can ask and answer the questions related to impact of specific community developments on citizen health.
- Spatio-temporal study of both disease and wellness of whole communities can be studied, and at the same time, differences in health outcomes of each citizen can be understood at a scale and granularity not possible today.

Desired Testbed Features and Related Challenges: An ideal testbed should allow capture of diverse forms of data from the participating citizens, and also enable novel interventions aimed to improve health and wellbeing. The key features of the testbed are:

- *City sensors*: Sensors that provide real-time information about factors which might influence health, e.g., air quality, transportation status, water quality and weather. While many such sensors are currently available and used, the real-time data is not readily made available. Further still, there are no application programming interfaces (APIs) to tap into the sensor data, and use it in novel methods and applications.
- *Participatory sensors*: Opt-in citizens that contribute data about their behaviors (activity, food consumption, sleep, stress), daily movement patterns (e.g., via their GPS), bio-markers and self-reported health outcomes. While many existing apps have been tested, seamless integration of all data modalities that can be deployed at large scales remains a significant challenge. Participatory sensors also pose unique challenges of data corruption, where an unintended user can use the sensor and corrupt the data. Thus, data provenance is a significant challenge in ensuring the utility of participatory data.
- *Security, privacy and fine-grained access control of data*: Each opt-in participant should have full (and dynamic) control of their data in regards who can view what part of their contributed data (including retrospective revocation of access). The key challenge is two-fold. First, any changes in access to data has to be propagated throughout the system. Second, the users

should understand the ramifications of their decisions, so that they can make informed choices.

- *Trusted data managers*: It is important that data managers are trusted by the community, and hence the natural managers of the sensitive personal information. Hospitals are one possible such trusted entities, since they already work with patients and have their sensitive data. However, for SC2 testbeds to be successful, it is important that the participants can choose their own trusted partner. This poses multiple challenges, including buy-in from many trusted healthcare providers, upgrading their infrastructure to support new forms of patient data mentioned above, and coordinating with other external partners to safely share the data.
- *Reconfigurable*: The testbed elements, city and participatory sensors, sensor modalities, data storage and access should be reconfigurable, allowing future growth and changes. Reconfigurability poses unique challenges between over-provisioning for future, and optimizing for cost and power.
- *Healthcare data*: For participating citizens, their healthcare data should be synchronized with their participatory data, allowing the study of health events and their potential causes. It will be important to get buy-in from primary, secondary and tertiary healthcare providers, with information both about their regular screenings and health events.

Existing Technologies and Future Work: Many recent developments could serve as a launch-pad for both development and deployment of community-scale testbed(s). On the engineering side:

- Apple HealthKit framework allows a large number of participants to enroll in IRB-approved clinical studies remotely, from their smartphones. Many studies have managed to enroll more than 100,000 participants in a short time.
- Many US cities have fixed ground infrastructure to measure air pollution or traffic conditions.
- Many app- and sensor-based mHealth studies, which capture contextual information, have already been conducted for targeted populations. They provide important learning on patient recruitment, retention, engagement, privacy and access control.

In terms of participant enrollment and existing communities with trusted data managers:

- US *Precision Medicine initiative* aims to recruit 1 million participants, making it one of the biggest planned prospective cohort study. The study is unique due to its size, and buy-in from different stakeholders. For example, 650,000 participants will be enrolled by the insurance providers and rest will be recruited via open enrollment.
- *SHARE*, Mediciti in Medchal Mandal, Hyderabad serves a rural population of 50,000+, has extensive information (both clinical and demographic) about every citizen and has already served as a clinical testbed for many important studies. For example, they have achieved nearly 100% immunization for all children and >95% hospital deliveries with a data-driven approach. Their experience will provide a blue-print for engaging a whole community and collecting many layers of information, beyond healthcare data.
- Agada Hospital in Chennai has a prospective cohort of 8000+ (1/3 urban, 1/3 semi-urban and 1/3 rural) to study diabetes and cardio-vascular diseases. The study is collecting extensive data on both bio-markers and behavioral patterns (e.g., food consumption). The study is unique in its diversity of enrollment from different socio-economic communities.

Collectively, the above technologies and community experiences will provide the much needed prior knowledge and experiences for successful large-scale SC2 testbeds.

Unique Opportunities for US-India Collaboration and Sharing Testbeds: It was identified that much of the required infrastructure for community-scale data collection, storage, ensuring privacy, security and access control can be co-developed and shared. The unique elements lie perhaps in human interfaces, e.g., for ensuring reliable participatory information in rural India may require different methodologies than in rural US. A joint testbed development effort could help identify both the commonalities and unique layers in development, thereby providing a principled approach in replicating such testbeds for new communities. The testbed research could itself become a template for actual community development.

5.2. Disaster Response and Recovery

Motivation for SC2 Testbeds on Disaster Response: The availability of large-scale testbeds was identified as a major need and enabler of research on disaster response, especially in SC2 contexts. Appropriate testbeds can contribute to the validation of new technologies developed in US-India collaboration efforts. On the data side, the workshop emphasized the importance of collecting joint data sets featuring both sensor and social data feeds about the same event. While social network data (e.g., Twitter and YouTube) is generally available via public APIs, sensor data that matches it is less widely available, but would be of great interest. Data collection efforts were encouraged that collect and clean social network and sensor data on main disasters to enable a suite of validation efforts aiming at investigating efficacy of different analytics, decision-support and situation understanding solutions. The collected datasets and testbeds should be aligned with research challenges and should distinguish rural versus urban environments. They should allow experimenting with different assumptions on solution cost and technology availability. Rural areas may benefit more from crowd-sensing solutions as well as solutions that rely on physical connectivity (e.g., the road network and physical data mules). In contrast, urban environments offer more opportunities for deployment of smart sensing infrastructure and exploitation of social networks. Availability of such testbeds and data sets will allow experimentation with the efficacy of developed technological capabilities in a range of realistic scenarios.

Desired Testbed Features and Related Challenges: Testbeds and datasets are needed to address three types of challenges:

- *Deployment time challenges:* When a disaster hits and response teams arrive to help with recovery, equipped with smart assets such as sensors and communication devices, several questions arise. For example, where and how to place smart assets? The solution may depend on where survivors are, where needs for sensing are, and how the nodes in question are to be connected. Algorithms are needed to deploy rescue and monitoring assets in an optimal fashion to gain situation awareness in a dynamic, evolving situation.
- *Run-time challenges:* Many key questions need to be addressed at run-time while recovery is in progress. For example, how to collect the “right” information? How to triage data based on human needs and value of information? How to leverage opportunistic information (e.g., information available on social media) to complement data collected from sensors? How to distinguish true and correct information from rumors? How to determine information reliability on social media? What algorithms are needed to facilitate decision making that balances inundating decision-makers with second-order details, versus not giving him enough relevant input? What are appropriate decision support tools that maximize first responder efficacy?
- *Design-time Challenges:* Several key questions must be addressed at system design time to reduce disaster recovery cost. For example, how to ensure robustness and resilience of infrastructure in the face of disasters? How to design adaptive systems that can be easily reconfigured to take advantage of resources left? How to understand and minimize propagation of failure cascades due to disaster events? How to model interactions between failures of different subsystems and the impact of such interactions on propagation of failure cascades? What are the implications on system architecture to maximize resilience and robustness?

Existing Technologies and Future Work: Many resources already exist that may be leveraged for experimentation. Testbeds include NIST wireless network experimental platforms (Boulder), landslide simulation testbed (Georgia Tech), vehicular traffic congestion simulators (e.g., for simulating efficacy of evacuation), and mesh network testbeds in multiple US cities. Datasets include repositories at government portals such as data.gov, as well as open portals for US smart cities, such as Chicago (including historical archives of past sensor data and surveys). Real-time information is also available in many cases, including traffic (e.g., California DOT has real-time data on traffic from all major highways). The US All Hazards Consortium offers datasets documenting damage from some major disasters. The US Geological Survey (USGS) has a comprehensive repository of weather related data. The CDC offers datasets on the spread of epidemics. Finally, real-time data may be collected from social networks such as Twitter using public programming APIs. These resources offer a good starting point for exploration and validation of disaster response technologies. Future work is needed to support experiments whose results can extend to SC2 and national scales.

Unique Opportunities for Indo-US Collaboration: An effective US-India collaboration can leverage the US testbeds and datasets, as well as expertise in analyzing and managing seasonal disasters, and technological artifacts such as self-driving cars, drones and robotics to use in disaster scenarios. India offers a unique environment with some of the highest world urbanization rates in terms of absolute population size converting to urban environments annually. New emerging SC2 are planned that offer an unprecedented opportunity for instrumentation at a large scale. Leveraging those resource will enable understanding the performance of tools at scales beyond what is currently possible. The collaboration will enable advances in disaster response in heavily populated and crowded environments.

5.3. Testbeds and Data Sets for Smart Energy and Buildings

Motivation for SC2 Testbed for Smart Energy and Buildings: The US and India have both made investments in university and industry environments to establish smart building and smart grid testbeds and pilot studies. The majority of these testbeds have been instrumented to capture the energy profile of building systems (photovoltaic (PV) arrays, heating, ventilation and air conditioning (HVAC), lighting, etc.) and occupancy behaviors. Examples of data gathered include building energy use aggregated by subsystem, type of load, geographic area, or occupancy level. Some testbeds also include auxiliary applications such as PV arrays, energy storage, and plug-in electric vehicles as well as other utilities (water and gas usage). The testbeds range in size from small micro grids in the 10s of kW range up to the MW scale of several buildings collectively.

Desired Testbed Features and Related Challenges: The workshop participants, in general, do not have ready access to the large datasets developed by other universities; some universities have started charging for access to their data. On the industrial side, datasets can typically be obtained from companies if requested, especially for Indian companies. However, datasets obtained from external sources may need to be reconfigured to be used and may not contain all of the desired vantage points.

Missing Gaps: One of the biggest gaps in the available testbeds and datasets is the lack of availability of large-scale, high resolution data. One of the challenges for establishing effective testbeds for smart energy is the expense of building and maintaining a useful testbed. For this reason, many of the smart energy testbeds are either large in scale but with low resolution or small in size but with high resolution. However, there is a significant need for large, fine-grained datasets. Another significant gap is the lack of availability of datasets that combine energy usage data and human activity, specifically with regard to behavior modification through information feedback or incentives. The majority of existing datasets contain large amounts of sensor data with little correlation to human activity that caused or resulted from changes in energy usage or environment.

Indo-US Sharing and Accessibility of Testbeds and Datasets: US and India have several complementary areas of expertise. There is considerable experience from both US and India researchers in energy storage applications. The US offers opportunities in energy storage at the grid scale (MW) whereas energy storage is distributed in India through widespread use of UPS systems. Both countries could benefit by a cross pollination of expertise as energy storage applications of all sizes become more prevalent. Furthermore, both countries have significant experience in the deployment of renewable energy sources. India provides a much greater opportunity for off-grid and micro grid applications whereas the US has a more probable outlook for a high penetration of plug-in electric vehicles.

There are several pertinent research avenues that naturally arise from the joint expertise of US and India researchers. One of the most compelling directions is to develop real-time monitoring and control systems for building energy management systems while including human in the loop activities. This encompasses the deployment of a non-invasive sensor network, adaptive controls that respond to human activities (both direct and indirect), and extension to a diverse population across economic, age, and occupational boundaries. The development of these complex systems provides an excellent opportunity to collaborate across multiple disciplines and universities in US and India.

Another relevant research direction is to exploit the widespread deployment of renewable resources and energy storage to develop energy management controls to better use the available resources. This approach can be developed for both the grid-connected (urban) setting as well as the off-grid (rural) micro grid setting. Appropriate energy management combined with distributed resources and energy storage can be used to provide improved reliability in all settings.

6. Best Practices and Lessons Learned in US-India Collaboration

This section is an overview of previous collaborations under US-India Pervasive Communications and Computing Collaboration (PC3) program launched in 2011. Significant results will be highlighted to demonstrate the value of collaborations in the sense that these outcomes would not have been possible by working in isolation. Finally lessons learned in PC3 projects will be discussed followed by best practices and mechanisms for successful international collaboration.

6.1. Past Collaborations in PC3 Program

The following five projects were jointly funded by the NSF and DeitY under the PC3 program.

- Smart Buildings (UCLA; IIIT Delhi, and industry partners)
- Smart Grids (UMass Amherst, IIT Bombay, and industry partners)
- Wireless Health (Dartmouth College, Rice University; IIT Delhi, AIIMS)
- Sensor Design for Wildlife Monitoring (Ohio State, Cornell, IISc Bangalore, IIIT Allahabad)
- Sensor Networks for Air Quality Monitoring (Univ. Rochester, Northeastern Univ.; IIT Delhi)

The PC3 program held three meetings: a preliminary workshop in New Delhi in March 2011, a kickoff meeting in Washington DC in June 2012, and a virtual PI/progress meeting in July 2013 involving the entire group of researchers as well as DeitY and NSF program officers.

6.2. Significant Results to Illustrate the Value of Collaborations

From a research standpoint, the collaborations provided an international context to research problems that were previously studied in local context (e.g., electric grid reliability issues in India are not common in the US context). They facilitated access to new international partners and new collaborations (beyond the institutions who started the project) which could not have occurred otherwise. Many projects led to research on frugal, portable and cost-effective designs leading to innovative potential product outcomes in some cases. Many datasets were created and shared between researchers and more broadly. For many researchers, the projects served as the seed for longer-term collaborations between Indian and US researchers that have outlasted the project durations and continue through this day. Finally, the projects led to successful student and faculty exchange and produced numerous joint research papers to major conferences and journals. The student exchange was especially successful since it provided US students with an international context and provided Indian students an opportunity to collaborate on top-tier research publications.

From an educational standpoint, an experimental joint course between a US and Indian educational institution was taught and it provided significant benefits to the students (such as joint student projects, internationalization experience; joint student mentoring).

There were many benefits from including industrial partners in the projects to both sides. Indian researchers were able to obtain access to datasets provided by US industrial partners, while US researchers could establish collaborations with Indian industry and research labs. Student internships at Indian companies (TCS, IBM India) provided an international industrial experience to students. The projects also resulted in collaborations with Indian hospitals and Indian non-profits (NGOs) with outcomes such as access to patient studies with more than 8000+ participants.

From translational research and commercialization standpoint, multiple projects produced intellectual property in the forms of patents. Three startup companies, two from the US and one from India, were spun-off based on work done as part of PC3 research.

6.3. Lessons Learned

Four lessons emerged out of the PC3 projects. First, forming the right team is crucial to the success of the project: early engagement between international partners can ensure a proper “match” and increases chances of a successful collaboration. Second, engagement with industry and non-academic partners significantly enriched the projects and provided real-world and international perspectives. Third, the project duration is key for certain types of projects---due to the challenges in establishing successful international partnerships, two years proved to be too short for certain types of PC3 projects (e.g., ones where there were task dependencies or ones that involved field work). A project duration of of 3-4 years may be more suitable for these cases. Fourth, different perspectives on common research problems enriched the various research projects.

6.4. Best Practices and Mechanisms to Support International Collaborations

Building on the PC3 experiences, best practices and mechanisms to ensure successful collaborations for SC2 projects include: (i) Student and faculty exchange in the form of short-term and semester-long visits, which greatly increase the breadth and depth of the collaboration; (ii) Including industry, governmental and non-academic partners, which is important to SC2 projects in domains such as healthcare and disaster management; (iii) Creation of shared datasets and larger shared SC2 testbeds, in the form of an entire smart city or smart village, which can provide an important focus to the projects; and (iv) Periodic project reviews in the form of semi-annual or annual workshops, which will ensure the continued success of the projects and facilitate face-to-face collaborations.

7. Conclusions and Recommendations

The SC2 workshop explored collaboration opportunities in the emerging topic of smart and connected communities, and addressed cutting-edge socio-technical research challenges in healthcare, energy and disaster response systems, in particular. The workshop produced a research agenda for developing innovative technologies, models, services and applications that can benefit citizens, urban planners, and policy makers in both countries. The bilateral cooperation between researchers in US and India has tremendous potential for unfolding the complexity of smart cities and connected communities in terms of healthcare, energy, and disaster response. This knowledge may have a disruptive effect on the organizational structure and development of a smarter and sustainable society. Given the enormous scientific and socio-economic benefits for both countries, we make the following recommendations regarding the collaboration between US and India on smart and connected communities.

- A collaborative funding program between US and India will be highly beneficial to the research and development of smart and connected communities in both countries. It will further lay the groundwork for long-term and broad collaboration between the two countries in areas beyond the focus areas identified at this workshop.
- We recommend selecting a model smart city and a smart village in each of US and India as open testbeds where the research outcome of funded collaborative projects will be deployed, integrated and evaluated. There is a need to provide funding and mechanisms for sharing, operating and maintaining testbeds and datasets across countries.
- A successful collaborative project will require at least three years due to the need for international coordination and field deployment. Furthermore, the timing of the US and Indian projects need to be synchronized to facilitate effective collaboration between teams in both countries.

Appendix A – List of Participants in Smart and Connected Communities Workshop

NSF Participants:

Kenneth Calvert, CISE/CNS Division Director
Harriet Taylor, CISE/CNS Program Director
Gurdip Singh, CISE/CNS Program Director

DeiTy Participants:

Brij Mohan Baveja
Tulika Pandey
Tara Shankar

US Organizers:

Sajal K. Das, Missouri Univ. of Science & Tech.
Chenyang Lu, Washington Univ. in St. Louis

Indian Organizers:

Uday Desai, IIT Hyderabad
Huzur Saran, IIT Delhi

US Academic Participants:

Tarek Abdelzaher (Univ of Illinois - Urbana)
R. Chandramouli (Stevens Inst Tech)
Mariesa Crow (Missouri S&T)
Sajal Das, (Missouri S&T)
Xiaofan (Fred) Jiang (Columbia Univ.)
Krishna Kant (Temple Univ.)
Santosh Kumar (Univ. of Memphis)
Chenyang Lu (Washington Univ. in St. Louis)
Calton Pu (Georgia Tech)
Ashutosh Sabhrawal (Rice University)
Prashant Shenoy (UMass-Amherst)
Mani Srivastava (UCLA)

Indian Academic Participants:

Laxminder Beheraq (IIT Kanpur)
Amrutur Bhardwaj (IISc Bangalore)
Debabrata Das (IIIT Bangalore)
Viswanath Gunturi (IIIT Delhi)
Devendra Jalihal (IIT Chennai)
Prem Kalra (IIT Delhi)
Kishore Kothapalli (IIIT Hyderabad)
Kiran Kuchi, (IIT Hyderabad)
Preetam Kumar (IIT Patna)
Rajiv Mishra (IIT Patna)
Sudipta Mukhopadhyay (IIT Kharagpur)
P. J. Narayanan (IIIT Hyderabad)
Rajalakshmi (IIT Hyderabad)
Vinay Ribeiro (IIT Delhi)
Anupam Shukla (IITM Gwalior)

Appendix B – Smart and Connected Communities Workshop Agenda

Day 1 (Thursday – June 9, 2016)

8:30 – 9:00: Registration

9:00 – 10:30: Opening Session (Chairs: Uday Desai and Sajal Das)

9:00 – 9:30: Lighting of Lamp

Welcome – Dr. Ram Gopal Rao (Guest of Honor), Director, IIT Delhi
Opening Remarks – Dr. Vijay Saraswat (Chief Guest), Member, NITI Aayog
Overview of Digital India – B.M. Baveja, Sr. Director and Group Coordinator, DeitY
NSF Welcome – Dr. Kenneth Calvert, NSF/CNS Division Director

9:30 – 10:00: Keynote Talk – Science and Technology for Smart Cities, Dr. Vijay Saraswat

10:00 – 10:30: Plenary Talks –

Welcome, and a Look Forward and Back – Dr. Jim Kurose, NSF CISE Assistant Director
Overview of NSF S&CC Activities – Dr. Gurdip Singh, CNS Program Director

10:30-11:00 – Tea Break

11:00 – 12:30: Self Introductions (Chairs: Huzur Saran and Chenyang Lu)

Follow the Quad Charts for Research Madness Presentations (3 mins each)

12:30 – 14:00: Lunch (IITD Faculty Guest House)

14:00 – 14:15: Format / Expectations of Breakout Sessions (Chairs: Huzur Saran, Sajal Das)

30 mins: (1) State of research in US (3-5 areas of active research related to SC2)
(2) State of research in India (3-5 areas of active research related to SC2)
(3) Unique strengths in US and India

30 mins: (4) 3-5 Missing gaps and research challenges to benefit from US-India collaboration
(5) 3-5 Opportunities for New Technology Development
(6) Open Questions to the whole group

30 mins: Summarize and prepare 5-6 slides using template for reporting to the whole group

14:15 – 15:45: Parallel Breakout Sessions on Healthcare, Energy, and Disaster Response

Healthcare (Lead – Amrutur Bhardawaj, Scribe – Santosh Kumar)
Energy (Lead – Fred Jiang, Scribe – Vinay Ribeiro)
Disaster Response (Lead – Kiran Kuchi, Scribe – Calton Pu)

Note: The main goal of this session is to brainstorm about (1) the state of research in each focused area in each country, and (2) top 5 or so research challenges as well as opportunities.

15:45 - 16:15 Tea Break

16:15 – 17:45: Presentations of Breakout Sessions (Lead/Scribe) & Preliminary Feedback

Healthcare (10 mins presentation + 10 mins Q&A)
Energy (10 mins presentation + 10 mins Q&A)
Disaster Response (10 mins presentation + 10 mins Q&A)
(30 mins) Discussions on Common Research/Technologies (Huzur Saran and Sajal Das)
Examples: Sensor Networks, CPS and IoT (Scribe – Mani Srivastava)
Wireless Communications (Scribe – R. Chandramouli)
Data Analytics (Scribe – Calton Pu)
Others?

Note: The lead/scribe of each focus area will present a summary of breakout discussions to all attendees and gather feedbacks; followed by open discussions on underlying technologies that potentially cut across multiple applications in the context of smart and connected communities.

17:45 – 18:15: Closing of Day 1 and Agenda for Day 2 (Chairs: Uday Desai, Chenyang Lu)

19:00 – 21:00: Banquet Dinner (IITD Main Guest House)

Day 2 (Friday – June 10, 2016)

8:30 – 8:45: Recap of Day 1 and Opening of Day 2 (Chairs: Huzur Saran and Sajal Das)

8:45 – 9:45am Indo-US Collaboration: Best Practices and Lessons Learned (Chenyang Lu)

(20 mins) Healthcare PC3 (Ashu Sabharwal)
(20 mins) Energy PC3 (Prasant Shenoy)
(20 mins) Shared Testbed Development (Uday Desai – Indo-Japan Experience)

Note: The invited talks will focus on experiences in Indo-US collaborations to set the stage for the next breakout session.

9:45 – 10:15: Tea Break

10:15 – 11:30: Parallel Breakout Sessions on Healthcare, Energy, and Disaster Response on Test-beds/Datasets/Associated Technologies

Healthcare (Lead – Ashutosh Sabharwal, Scribe – Anupam Shukla)
Energy (Lead – Debabrata Das, Scribe – Mariesa Crow)
Disaster Response (Lead – Rajiv Mishra, Scribe – Tarek Abdelzaher)

Note: The discussion will focus on testbeds and data sets that can be exploited, developed and shared through Indo-US collaboration on the focus areas.

11:30 – 12:45: Summaries of Test beds, Datasets, and Technologies (Lead/Scribe)

Healthcare (10 mins presentation + 5 mins Q&A)
Energy (10 mins presentation + 5 mins Q&A)
Disaster Response (10 mins presentation + 5 mins Q&A)
(30 mins) Discussions on Sharing and System Integration of Test beds and Data Sets

12:45 – 13:45pm: Lunch

13:45 – 14:45: "Why US-India Collaboration – Benefits and Mechanisms?" (*Lead: Krishna, Scribe: Prashant Shenoy*)

Note: this session will be for open discussions on best practices and mechanisms for Indo-US collaboration and sharing.

14:45 – 15:45: Topical Breakout Discussions to Organize Thoughts and Start Writing – Develop Framework, Outline for Main Ideas to Discuss in Writing, Assign Sections to Group Members

Note: The goal of this session is to set up the outline and structure of the section on each application in the workshop report; the writing lead of each section will further assign subsections to different attendees; the group will then start adding the key points to the subsections in parallel.

15:45 – 16:15: Tea Break

16:15 – 17:15: Parallel Writing Sessions (Writing Leads and Associates)

17:15 – 17:45: Writing Preview (Lead/Scribe) and Critical Feedback

(10 mins) Smart Healthcare
(10 mins) Smart Energy
(10 mins) Disaster Response

Note: The writing lead of each section will go over the current version of the section to all attendees for early feedbacks.

17:45-18:00 Closing Remarks of Day 2 and Agenda for Day 3

18:00 – 22:00: Cultural Exchange – Musical Concert (India Habitat Centre) and Dinner

Day 3 (Saturday – June 11, 2016)

9:00 – 9:15: Recap of Day 2 (Chairs: Chenyang Lu and Kiran Kuchi)

9:15 – 10:45: Parallel Writing Sessions

10:45 – 11:15: Tea Break

11:15 – 12:30: Writing, Presentations and Discussions on Integrated Report

Note: The writing lead of each section will go over the current version of the section to all attendees, followed by discussions on the integrated report.

12:30 – 14:00: Lunch

14:00 – 19:00: Organizers Finalize the Report and Free Time for Others

Note: The goal is for organizers to finish the report before departure.