





Prepared by:

Prasad Calyam (Chair), University of Missouri-Columbia & Glenn Ricart (Co-Chair), US Ignite and U. Utah

Acknowledgements

Support for this workshop was provided by the National Science Foundation¹ and we are grateful for the helpful guidance of NSF Program Manager John Brassil in the workshop organization. Special thanks to the NSF "Looking Beyond the Internet" planning group led by Chip Elliott, Dave Farber and Larry Landweber for providing a vision and resources for a series of three workshops (including this workshop) that are intended to uncover new research opportunities in the broad area of potentially integrated infrastructure for the year 2021 Internet, Cloud, and Wireless/Mobile domains. We would also like to acknowledge the report contributions of the workshop track discussion leaders: Nishal Mohan, Micah Beck, Malathi Veeraraghavan, Rick McGeer, Sean Smith and Kaiqi Xiong. Lastly, we would like to sincerely thank Jeffrey Blumer, University of Missouri-Columbia for the administrative support in the workshop co-ordination and travel grant management.



¹ This material is based upon work supported by the National Science Foundation under Award Number CNS-1546769. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Invited Participants

Ann DeVilbiss, One Community	Marlon Pierce, Indiana University
Arnold Glass, Rutgers University	Micah Beck, University of Tennessee at Knoxville
Balakrishnan Prabhakaran, University of Texas at Dallas	Mina Sartipi, University of Tennessee at Chattanooga
Bruce Patterson, City of Ammon	Nishal Mohan, US Ignite
Christos Papadopoulos, Colorado State University	Parmesh Ramanathan, University of Wisconsin-Madison
Comiter Marcus, Harvard University	Patricia Campbell, KINBER
Cynthia Chen, University of Washington	Paul Ruth, RENCI
David Lary, University of Texas at Dallas	Kapil Dandekar, Drexel University
David Matolak, University of South Carolina	Raj Jain, Washington University in StLouis
Deborah Estrin, Cornell Tech	Raj Kettimuthu, Argonne National Laboratory
Florence Hudson, Internet2	Arcot Rajasekar, RENCI
Houbing Song, West Virginia University	Rick McGeer, US Ignite
Ibrahim Demir, University of Iowa	Rudra Dutta, North Carolina State University
John Day, Boston University	Rui Dai, University of Cincinnati
Jason Liu, Florida International University	Sean Smith, Dartmouth College
Jeff Jaffe, W3C	Shivakant Mishra, University of Colorado Boulder
Joe Mambretti, Northwestern University	Steven Corbato, Oregon HSU
Kaiqi Xiong, University of South Florida	Suman Banerjee, University of Wisconsin-Madison
Kate Keahey, University of Chicago	Tilman Wolf, University of Massachusetts-Amherst
Kwa-Sur Tam, Virginia Tech	Ting Zhu, University of Maryland-Baltimore County
Larry Smarr, UC San Diego	Tracy Futhey, Duke University
Leon Gommans, University of Amsterdam	Vetria Byrd, Purdue University
Lev Gonick, One Community	Venky Krishnaswamy, Avaya Labs
Malathi Veeraraghavan, University of Virginia	Weisong Shi, Wayne State University
Marilyn Wolf, Georgia Tech	Yan Luo, University of Massachusetts Lowell
Marjorie Skubic, University of Missouri	Yu Wang, University of North Carolina at Charlotte

























1. Executive Summary

Today's digital infrastructure is undergoing deep technological changes and new paradigms are rapidly taking shape in both the core and edge domains. These paradigms leverage the growing footprint of ultrahigh-speed broadband networks, pervasive wireless, cloud computing, and software-defined infrastructure. Moreover, they are positioned to connect smart/mobile devices, as well as their data on a massive scale. These advances will enable transformative applications and services in the decade beginning in 2021 that will enhance the quality of peoples' lives while addressing important national priorities. In addition, these new technologies may bridge and unite today's cutting edge applications such as the Internet of Things, Big Data Analytics, Robotics, The Industrial Internet, and Immersive Virtual/Augmented Reality, yielding applications that are not obvious or even possible at this time.

The NSF-sponsored 'Applications and Services in the Year 2021' workshop was successfully organized on January 27-28, 2016 in Washington DC that brought together applications researchers in multidisciplinary areas, and developers/operators of research infrastructures at both national, regional, university and city levels. Discussions were organized to identify grand challenge applications and obtain the community voice and consensus on the key issues relating to applications and services that might be delivered by advanced infrastructures in the decade beginning in 2020. As such the topics discussed, the findings from the workshop, and the recommendations for future NSF funding areas can be classified under "Research" needs, and "Infrastructure" needs as follows.

Research

Finding-1) An explosion of data, especially from communicating devices, will upset architectures and demand new approaches. Data will be at the center of much that drives future research and development, and the economy. It will be the glue that unites the physical and cyber worlds, the key to understanding human environments, the intelligence behind personal cyber coaches, the comparative weight that allows us to balance privacy against security in each case, and the grounding for artificially intelligent agents acting on our behalf.

- **Recommendation-1:** Improved data science techniques for harvesting, collecting, processing, compressing, storing/archiving, sharing, and transmitting data are needed to deal with the data <u>'deluge'</u> we (e.g., researchers, enterprises, and cities) are increasingly experiencing. Archived information from a variety of data sources will be the basis for groundbreaking research, and hence techniques to make data more accessible and manageable are critical.
- Recommendation-2: "Edge computing" paradigms are needed that are secure, privacy preserving, and able to support real-time analysis in multi-tier infrastructure environments, in spite of any first-mile (data collection sites) and last-mile (data consumption sites) network limitations. Research on Locavore (edge) clouds (also called Fogs and Cloudlets) is urgently needed to explore how they can best accept, learn from, and take decisions based on massive streams of local IoT data, while also sharing contextual information with other locavore clouds.

Finding-2) Applications in the national priority areas (healthcare, education, public safety, citizen innovation, etc.) are good drivers for interdisciplinary research. An especially good example is disaster management because it pushes science in many areas, and requires impactful interdisciplinary collaborations. Support for science advances and innovations in national priority areas can benefit other application areas or wider technology infrastructure in rural areas or under-developed country regions, and have broad societal impact.

- **Recommendation-1:** Interdisciplinary collaboration should be strongly encouraged because it will be a foundational aspect for addressing Global Challenges (approaching as a community rather than as separate disciplines). A "systems approach" is needed with more of an interdisciplinary team approach (machine learning+ videoanalysis+ sensors+ networking+ HPC+ SDN+ security+ datascience+ healthscience+ policy+....) i.e., "diversity will breed bold discovery".
- **Recommendation-2:** <u>Research and development for disaster management applications and new</u> <u>techniques for related infrastructure setup and management should be particularly encouraged</u>. This is because advances for this example application will have to address multi-disciplinary challenges and overcome the fact that general assumptions of infrastructure availability may not hold true for users in a disaster incident area.

Finding-3) An interesting part of the applications space is its growing need for "smart services" for continuously operating ecosystems of data, interactive analysis, artificial intelligence interpretation, and human-guided visualization and intervention. There is a need to better engage human as well as artificial intelligence to gain 'insights' from data that can then be used to design better applications and adapt them based on user needs.

- **Recommendation-1:** <u>There is an urgent need for innovation in 'interactive' and 'collaborative'</u> <u>visualization services to obtain relevant insights from the data</u>. Visualization needs to leverage ultra-low-latency links where available, ad hoc visualization of massive datasets (immersive reality, head-mounted displays) and be seamlessly accessible across multiple form factors (e.g., mobile device versus immersive displays).
- **Recommendation**-2: <u>There is a need for research into "Insights-as-a-Service" that operates as an integrated function i.e., it jointly considers function-centric orchestration in data movement, machine learning, data sharing and computing for 'ubiquitous analytics purposes'. Such an Insights-as-a-Service paradigm can help us handle the fact that data and processing will be increasingly distributed in nature and (sometimes) federated.</u>

Finding-4) *Privacy and security is ill-understood in this new world (with IoT growth) both technically and socio-technically and also socio-economically*. There are many unsolved hard problems in today's users managing privacy when they interact with web servers over the "Internet of Computers" (IoC), and thus it will be much harder in the "Internet of Things" (IoT) future for users to express their privacy wishes and understand their implications in context of ubiquitous sensors and invisible applications.

- **Recommendation-1:** <u>Innovative techniques and usable tools for composing privacy policies and managing privacy preferences need to be developed</u>. In the multipoint/reuse infrastructure in 2021, user data will flow across many boundaries and merge and split and aggregate. Novel research and technology should be supported that can help users or communities to compose privacy policies that make sense in a multipoint/reuse infrastructure and also carry privacy as per user/community desire throughout the data lifecycle.
- **Recommendation-2:** <u>At-scale research experimentation should be encouraged with establishment of large testbeds of real users with diverse preferences in order to create a privacy-protected repository of preference data that can be used/re-used by researchers. Academic research on data privacy should recruit subjects with life experience or broader perspective to reflect the diversity and complexity of the population at large to raise logistical and (ironically) privacy challenges that can be addressed through large-scale user/infrastructure experimentation.</u>

Infrastructure

Finding-5) Cities, enterprises and university citizen communities can be leveraged as interesting initial testbeds and even 'living labs' to solve the application and service challenges in a IoT world of the year 2021. The IoT infrastructure supporting the applications of 2021 will create an even more complex tangle of data and computational channels—tiny devices and big-data backends—spanning jurisdictions, economies and cultures.

- **Recommendation-1:** <u>Students and citizens should be encouraged to "live in the future" as part of</u> <u>'living labs', and push the limits of what we can "imagine" in terms of futuristic possibilities, and</u> <u>unexpected new frontiers for innovation</u>. Such efforts should focus on improving the productivity as well as improving quality of human life through the use of new and advanced technologies.
- **Recommendation-2:** Living lab testbeds can foster interdisciplinary research, marrying the IT fields of computer science and engineering with the human-oriented fields of sociology and psychology and economics. In addition, specific "grand challenges" can be pursued; one might study of culture and trust amongst users in smart city eco-systems, where research experiments enable "home testing" of security/privacy in the IoT infrastructure of 2021 (the same way users can test their tap water today). Alternately, these testbeds can help overcome any regulatory obstacles in public spaces that have been a significant barrier to deployment of IoT and cyberphysical systems (e.g., in case of UAVs or wearable computing in public spaces).

Finding-6) *Infrastructure that is broadly available should allow for testing of Year 2021 applications and services at scale*. Driven by the data explosion as well as the consumerization of devices that brings more and more people online—not to mention devices/IoT—systems and applications need to be designed to quickly expand and scale.

- **Recommendation-1:** There is need for federated, deeply programmable, open, sliceable and sharable infrastructure for experimentation (similar to current GENI, CloudLab, Chameleon and beyond). Locavore infrastructure is especially needed at the edge for applications with low-latency requirements. With broadly available infrastructure that can be operated as software-defined infrastructure, new solutions for national priority application areas (such as smart manufacturing systems) can be realized.
- **Recommendation-2:** Infrastructure should be designed and deployed in simple abstractions as well as with realistic application targets and use cases. Proposed approaches for infrastructure design should have minimal specificity (end-to-end), generic, and simple enough to maximize the generality of the resulting solution across multiple disciplines and user groups.

Finding-7) The infrastructure supporting applications and services in the year 2021 will take the form of an IoT that has unsolved security challenges due to: massive numbers of small devices, or devices with increasing data resolutions (relatively larger devices), each engineered to become inexpensive, massively distributed throughout physical space over time and more widely shared. Several factors such as: scale of devices, size of data sets, economy due to cheaper manufacturing methods, increased distribution and sharing—will all conspire to create new security challenges.

• **Recommendation**-1: <u>New techniques and related tools to make it easier for researchers and</u> programmers to build distributed systems the right way quickly from a security standpoint should <u>be developed</u>. The infrastructure of 2021 will require even more distributed systems that are quickly. Flexible and extensible web technology stack research is needed to handle the technical, policy and legal specifications relating to infrastructure resources owned by multiple entities servicing users/devices at massive scales.

• Recommendation-2: Investigation of innovative methods should be encouraged that allow automated security verification for human analysts and software engineers. Proposed solutions would likely need to balance the advantage of full white-box testing with the privacy of source and design internals; successful solutions here will also likely need to figure out how to significantly increase assurance that this system-of-systems is secure, while recognizing the fundamental difficulty of "100%" confidence.

2. Workshop Organization Process and Methodology

2.1 Topics and Participants Selection

The timing and organization for this workshop is significant and well summarized by the "Looking Beyond the Internet" Planning Group Organizers – "Now is the time to make concrete plans for exploring emerging planetary-scale cloud / wireless / network systems beyond the Internet …".

The following is a list of themes that were identified for setting the scope of the workshop discussions and to relay the topics of interest:

- User quality of life applications: use cases
- Next-generation applications and services
- Application run-time environments on federated and cloud resources
- Real-time data collection, analytics and sharing
- Supporting infrastructure including the roles of clouds, wireless/mobile and SDI technologies
- Federation of infrastructure for supporting applications and services
- Security and privacy requirements of applications and services in the decade beginning 2021

An open invitation was sent out to the research community and some industry contacts to submit onepage whitepapers that included: (i) a brief bio, (ii) a set of research challenges and opportunities relating to the Workshop's themes that should be discussed in the workshop. A workshop website was also created – <u>http://asw2016.wordpress.com</u>

More than 90 whitepaper submissions were received, out of which 58 participants were invited to the workshop. A select number of participants were awarded travel grants based on need and availability of funds. The invitations were based on creating a reasonable mix of researchers, industry, city managers, regional networks and PhD students to have a diverse set of perspectives in the discussions.

There were many interesting "Imagine a world" and "Imagine a time/place" stories provided in the whitepapers. A couple of examples are as follows:

- "We envision a time when sensor technology and smart healthcare systems are routinely used across the U.S. to capture subtle changes in health that may be early signs of illness and functional decline. Seniors are able to receive early treatment when health problems are still small and manageable; healthcare is more effective and efficient, because health problems are caught early..."
 - Marjorie Skubic, University of Missouri-Columbia

- "A research team has been awarded time on an expensive instrument. With a few clicks, they establish a collaboration space containing data relevant to the specific experiment. Several members of the research team plus instrument scientists work within this space to prepare for the experiment. The sample(s) arrive ... The entire process happens in an interactive fashion."
 - Rajkumar Kettimuthu, Argonne National Laboratory

Other examples included references to: Smart Car, Smart City, Smart Grids, Smart Industrial Systems, Smart Flood Information Systems, Smart Disaster Management, and Smart Education Technology.

The other highlights from the accepted workshop whitepapers include:

- Application Areas Represented
 - Education, Public Safety, Civil Engineering, Transportation, mHealth, Eldercare, Earth Sciences, Energy, Genomics, Clinical Imaging, Manufacturing, ...
- Smart Technologies Represented
 - IoT: UAVs, Wearables, Sensors, Virtual Reality, Augmented Reality, Robotics, 3D immersion, Haptics, ...
- Data Intelligence Representation
 - Small Data, Big Data, Federated Data, Crowd-sourced Data, Multi-sensoral Media
- Infrastructure and Services Representation
 - Software-defined Infrastructure and Services, Edge/Core Cloud, Hybrid Cloud, Hierarchical Clouds, Named Data Networking, Science Gateways, Sustainable App Ecosystems, Visualization, ...
- Security and Privacy

The workshop participants were sent a pre-workshop reading list that included:

- 2014 Report of the NSF CISE AC Midscale Infrastructure Committee
- 2002 NSF Workshop Report on Network Research Testbeds
- 2015 NSF-funded Cybersecurity Experimentation of the Future (CEF) Report
- Whitepapers of Invited Participants

2.2 Workshop Discussion Themes

The main workshop goals were as follows:

- 1. Discuss several "themes" of research challenges in group discussions
- 2. Identify a set of "salient findings" and "recommendations" for 'Applications and Services in the Year 2021'
- 3. Synthesize 'community voice/consensus' in a final report to NSF

Below table shows sample keywords announced by participants during the workshop kickoff session. These keywords set the tone for rest of the workshop discussions that had diverse perspectives and interesting topics relevant to applications and services in the year 2021.

Good Science First	Data infrastructure intelligence
Home-based Healthcare	 Visualization facilitates insight
• Dark Factory	Multi-point Accountable/Reusable Apps
Quality of Information	• Inventing the Inter-not

- Interactive data analytics
- Edge Computing infrastructure
- Security/Privacy Everywhere
- Software-defined Smartphones
- Machine Learning with Networks
- Cloud Service Continuity
- High Performance Simplified
- •

2.3 Workshop Program

Wednesday, Jan 27th

- Plenary Session #1 (8:30am 10:30am)
 - NSF Welcome (10 Minutes)
 - Workshop Goals/Agenda Review: Prasad Calyam (20 Minutes)
 - Chair Talk: Glenn Ricart (30 Minutes)
 - Keynote Speech #1: Tracy Futhey Sharing Economy influence on Future Apps (30 Minutes)
 - Short Talk #1: Florence Hudson Policy & Apps Tech (15 Minutes)
 - Short Talk #2: Suman Banerjee Edge Computing for Apps (15 Minutes)
- Break (10:30am 11am)
- Plenary Session #2 (11am 12pm)
 - Keynote Speech #2: Raj Jain Future IoT/Apps Protocol Stack (25 Minutes)
 - Short Talk #3: Nishal Mohan Apps Ecosystem (15 Minutes)
 - Short Talk #4: Sean Smith IoT Security (15 Minutes)
 - Breakout Sessions Logistics: Prasad Calyam (5 Minutes)
- Lunch (12pm 1pm)
- Breakout Sessions (1pm 3pm)
 - Format: 4 Tracks with Scribe Leads assigned, Participant Short-talks in each Track (5 minutes duration per talk) along with Group Discussion
- Coffee Break (3pm 4pm)
- Breakout Session Reports (4pm 5pm)

Thursday, Jan 28th

- Breakfast (7:30am 8:30am)
- Combined Session (8:30am 10:00am)
 - Keynote Speech #3: Larry Smarr (45 Minutes)
 - Keynote Speech #4: Deborah Estrin Small-data fueled Apps and Services (25 Minutes)
 - Short Talk #5: Marjorie Skubic Health Tech & Apps
 - Combined Session Logistics: Prasad Calyam (5 Minutes)
- Break (10:00am 10:30am)
- Group Discussions (10:30am 11:30am)
 - Research Agendas for 4 Tracks
 - Testbeds for 4 Tracks
- Group Discussion Reports (11:30am 12pm)
- Closing Remarks (Prasad Calyam and Glenn Ricart)
- Box Lunch (12:15pm 1pm)
- PC Report Writing (1pm 4pm) (Chairs and Scribe Leads)

9

- Open/Virtual Infrastructure
- Realtime Healthcare
- Named Data Networking
- Security/Privacy Attack Defenses
- Proactive healthcare
- Fallacy of security/privacy
- Explosion of Experimental Sciences
- ...

The workshop breakout discussions on the first day were organized under 4 separate tracks:

- Track-1: Applications Grand Challenges
 - Discussion Leader: Nishal Mohan; Supporting Scribe: Glenn Ricart
 - Short Talks: Nishal Mohan, Mina Sartipi, Ibrahim Demir, Lev Gonick

• Track-2: Data Intelligence and Visualization

- Discussion Leader: Micah Beck; Supporting Scribe: Prasad Calyam
- Short Talks: Micah Beck, Arcot Rajasekar, Balakrishnan Prabhakaran, Comiter Marcus

• Track-3: Services and Infrastructure

- Discussion Leader: Malathi Veeraraghavan; Supporting Scribe: Rick McGeer
- Short Talks: Malathi Veeraraghavan, Jason Liu, Raj Kettimuthu, Leon Gommans

• Track-4: Security and Privacy

- Discussion Leader: Sean Smith; Supporting Scribe: Kaiqi Xiong
- Short Talks: Sean Smith, Kaiqi Xiong, Rudra Dutta, Shivakant Mishra

To ensure diversity of opinion and expertise in groups, and balance the spread of participants, we suggested tracks during the breakout sessions for participants to attend. Each track had a working document that was open for all participants to add content. Track Chairs synthesized these documents into a separate document for 'breakout session' and 'combined session' reports.

On the second day, a combined track was organized for group discussions that formed a "Take-Two" of the Day-1 breakout tracks. The goal here was to increase the interactions of the various breakout groups, and thus we decided to mix up the participants in various discussion areas setup at 4 corners of the main conference room. Discussion leaders from Day-1 breakout sessions coordinated the group discussion at each of the respective track corners. There were 20-minute interruption announcements, so that participants could walk over to the other track discussion corners one after the other. Participants were assigned to 3 discussion corners that had tracks they did NOT attend on Day-1. e.g., Track-1 participants on Day-1 went to Track-2, 3 and 4 corners.

The presentation slides of the workshop speakers as well as the questions that guided the group discussions (also listed in **Appendix-I** of this report) are available on the workshop website – <u>http://asw2016.wordpress.com</u>

A list of Application Grand Challenges that were identified by Track-1 group discussions are separately included in **Appendix-II** of this report.

3. Workshop Findings and Recommendations

Research

Finding-1) An explosion of data, especially from communicating devices, will upset architectures and demand new approaches. Data will be at the center of much that drives future research and development, and the economy. It will be the glue that unites the physical and cyber worlds, the key to understanding human environments, the intelligence behind personal cyber coaches, the comparative weight that allows us to balance privacy against security in each case, and the grounding for artificially intelligent agents acting on our behalf. There will be immense amounts and many types of data with increasing volumes that will need to be organized and processed. Janusz Bryzek at the TSensors Summit suggested that one trillion devices will be communicating in our target year of 2021. These numbers suggest our current architectures designed for managing data of e.g., web pages for humans may not be appropriate under this level of machine-to-machine communications.

- Recommendation-1: Improved data science techniques for harvesting, collecting, processing, compressing, storing/archiving, sharing, and transmitting data are needed to deal with the data <u>'deluge'</u> we (e.g., researchers, enterprises, cities) are increasingly experiencing. Archived information from a variety of data sources will be the basis for groundbreaking research, and hence techniques to make data more accessible and manageable are critical. As fast as the use of computer cycles grows, as fast as the gigabit networks grow, nothing will grow faster than data being acquired by devices. Envisioning reasonable architectures for triaging the data and storing appropriate, quality-tagged, ontologically-organized, and useful-life-dated data is a significant research area. Effective data triaging, integration and sharing from various sources and varied sensors through future (e.g., self-organizing, self-governing, software-defined) data architectures could help solve grand challenges such as 'planet stewardship', where we are able to more effectively monitor and dynamically model our climate and environment to take necessary actions to limit negative impacts for our future generations.
- Recommendation-2: "Edge computing" paradigms are needed that are secure, privacy preserving, and able to support real-time analysis in multi-tier infrastructure environments, in spite of any first-mile (data collection sites) and last-mile (data consumption sites) network limitations. Research on Locavore (edge) clouds (also called Fogs and Cloudlets) is urgently needed to explore how they can best accept, learn from, and take decisions based on massive streams of local IoT data, while also sharing contextual information with other locavore clouds. They provide a middleground between the cloud and local computing in terms of its relatively medium-latency and medium-throughput, which can provide interesting benefits to future applications and services. In addition, research should be encouraged in edge-computing to investigate issues that include ensuring symmetric gigabit access, ultra-low latency, seamless edge/core compute cycle access, pricing methods, suitable service interfaces, usability, and security tools for device-to-service or device-to-device communications. It is important to note that research investments in fog/edge computing for the creation of large-scale distributed systems could counter-balance the current commercial forces driving cloud computing.

Finding-2) Applications in the national priority areas (healthcare, education, public safety, citizen innovation, etc.) are good drivers for interdisciplinary research. An especially good example is disaster management because it pushes science in many areas, and requires impactful interdisciplinary

collaborations. Support for science advances and innovations in national priority areas can benefit other application areas or wider technology infrastructure in rural areas or under-developed country regions, and have broad societal impact.

- **Recommendation-1:** Interdisciplinary collaboration should be strongly encouraged because it will be a foundational aspect for addressing Global Challenges (approaching as a community rather than as separate disciplines). A "systems approach" may likely drive a need for future research agendas to be much broader and with more of an interdisciplinary team approach (machine learning+ videoanalysis+ sensors+ networking+ HPC+ SDN+ security+ datascience+ healthscience+ policy+....) i.e., "diversity will breed bold discovery".
- Recommendation-2: <u>Research and development for disaster management applications and new</u> techniques for related infrastructure setup and management should be particularly encouraged. This is because advances for this example application will have to address multi-disciplinary challenges and overcome the fact that general assumptions of infrastructure availability may not hold true for users in a disaster incident area. Research investigations that leverage the new software-defined networking (SDN) and network function virtualization (NFV) paradigms have the potential to overcome disaster management barriers caused by the dominance of the deployed Internet infrastructure and protocol stack. Although there has been research literature on 'Disruption Tolerant Networks' to handle the resilient network needs of applications, they have not considered the massive data handling use cases. Research to deal with data management for disaster management could result in fulfilling data management needs of other Big Data applications under adverse situations or with corrupt or missing infrastructure or metadata.

Finding-3) An interesting part of the applications space is its growing need for "smart services" for continuously operating ecosystems of data, interactive analysis, artificial intelligence interpretation, and human-guided visualization and intervention. Tools in sophisticated statistical software have long had poor graphing that missed many insights on data; we are already collecting and will increasingly collect huge amounts of data for which new statistical analysis techniques are being proposed, yet visualizations (e.g., in existing statistical software packages) are poor in terms of providing pertinent insights. There is thus a need to better engage human as well as artificial intelligence to gain 'insights' from data that can then be used to design better applications and adapt them based on user needs.

- Recommendation-1: There is an urgent need for innovation in 'interactive' and 'collaborative' visualization services to obtain relevant insights from the data. Visualization needs to leverage ultra-low-latency links where available, ad hoc visualization of massive datasets (immersive reality, head-mounted displays) and be seamlessly accessible across multiple form factors (e.g., mobile device versus immersive displays). It also has to accommodate real-time as well as opportunistic (context-aware) mechanisms to analyze data (versus today's mostly offline data processing and subsequent visualization) using interactive work involving humans or services. It would be useful to investigate sensor profile related analytics and visualization approaches, which provide end-users with a suite of algorithms that rapidly adapt to changing priorities and needs in exploring massive datasets.
- Recommendation-2: There is a need for research into "Insights-as-a-Service" that operates as an integrated function i.e., it jointly considers function-centric orchestration in data movement, machine learning, data sharing and computing for 'ubiquitous analytics purposes'. Such an Insights-as-a-Service paradigm can help us handle the fact that data and processing will be

increasingly distributed in nature and (sometimes) federated. To realize an effective Insights-as-a-Service capability, there is a need for the sharing/interoperability in distributed systems to extend beyond just networking; there should be seamless sharing/interoperability across other components such as storage and computation services within future fog (edge-cloud) and corecloud infrastructures. A real benefit of Insights-of-a-Service can be argued to meet the needs of the Industrial Internet that has huge raising research problems in information fusion (remote sensing, sensor networks, social media data streams), dealing with rich information from hyperspectral sensors, aggregating massive amounts of "small data," and deriving knowledge and insights from the collective awareness of autonomous things.

Finding-4) *Privacy and security is ill-understood in this new world (with IoT growth) both technically and socio-technically and also socio-economically.* There are many unsolved hard problems in today's users managing privacy when they interact with web servers over the "Internet of Computers" (IoC), and thus it will be much harder in the "Internet of Things" (IoT) future for users to express their privacy wishes and understand their implications in context of ubiquitous sensors and invisible applications. There will be greater challenges in user consent, scenarios where there user/server interaction involves multiple parties with their own vested interests, and enforcement of user preferences as data is used and propagated amongst multiple parties for social or financial purposes.

- Recommendation-1: Innovative techniques and usable tools for composing privacy policies and managing privacy preferences need to be developed. In the multipoint/reuse infrastructure in 2021, user data will flow across many boundaries and merge and split and aggregate. Novel research and technology should be supported that can help users to compose privacy policies that make sense in a multipoint/reuse infrastructure and also carry privacy as per user desire throughout the data lifecycle. If the ubiquitous IoT applications of 2021 are to respect user privacy preferences, then the technology has to effectively embody these preferences; put more succinctly, these concepts have to cross the barrier from me-space to cyberspace and be suitable in multi-domain network/distributed system infrastructures. However, we should be careful in giving users machines with larger and more complicated sets of knobs; such an approach will not work when the machines and their data are invisible in an IoT world.
- Recommendation-2: <u>At-scale research experimentation should be encouraged with establishment of large testbeds of real users with diverse preferences in order to create a privacy-protected repository of preference data that can be used/re-used by researchers. Academic research on data privacy too often uses small groups of e.g., university students or citizen community as the test subjects. Although useful results may be obtained, such subjects will not have the life experience or broader perspective to reflect the diversity and complexity of the population at large. (E.g., can healthy 22-year-olds who've never had to worry about finding and paying for health insurance even without pre-existing conditions have anything meaningful to contribute about health data privacy?) Establishing larger and diverse subject populations will raise logistical and (ironically) privacy challenges that can be addressed through innovative research that is guided by experimentation at realistically large scales of users and infrastructures.</u>

Infrastructure

Finding-5) Cities, enterprises and university citizen communities can be leveraged as interesting initial testbeds and even 'living labs' to solve the application and service challenges in a IoT world of year 2021. The IoT infrastructure supporting the applications of 2021 will create an even more complex tangle of data and computational channels—tiny devices and big-data backends—spanning jurisdictions, economies and cultures. Deploying applications and services in this new world will require understanding not just the technology, but also the human individuals and societies actively using them for continuous improvement.

- Recommendation-1: Students and citizens should be encouraged to "live in the future" as part of 'living labs', and push the limits of what we can "imagine" in terms of futuristic possibilities, and unexpected new frontiers for innovation. Such efforts should focus on improving the productivity as well as improving quality of human life through the use of new and advanced technologies. Research can be targeted to study how the integration of human assistive technologies such as voice assistance and cloud computing could create many new applications for user convenience and productivity. In addition, it is important to provide students/citizens access to the long tail of unpublished data, failed, or one-off experimental data of researchers. This could promote new ways of thinking amongst students/citizens and may be vitally important to address scientific reproducibility and also put a spotlight on data quality. Similarly, an overlying infrastructure for the deposit, storage, search, and use of deployable execution environments, particularly including Virtual Machines and containers, as well as executable computational science workflows (based on existing and new workflow engines) can be setup to enhance reproducibility of experiments.
- **Recommendation-2:** Living lab testbeds can foster interdisciplinary research, marrying the IT fields of computer science and engineering with the human-oriented fields of sociology and psychology and economics. In addition, specific "grand challenges" can be pursued; one might study of culture and trust amongst users in smart city eco-systems, where research experiments enable "home testing" of security/privacy in the IoT infrastructure of 2021 (the same way users can test their tap water today). This will help us better understand the dependence of security and privacy on user behavior and culture. For e.g., it helps us to reason about which higher goals justify sacrificing privacy protection. Divergence of user demographics also leads to divergent trust attitudes about potential validators and enforcers of the security and privacy properties of technology. Alternately, these testbeds can help overcome any regulatory obstacles in public spaces that have been a significant barrier to deployment of IoT and cyberphysical systems (e.g., in case of UAVs or wearable computing in public spaces). Proper institutional review board oversight should be sought to support such research, and suitable incentives for user opt-in should be explored. Further, these 'living lab' experimental facilities when jointly shared by industry and academic researcher teams could lead to new innovations that can be rapidly translated to benefit enterprises and even citizen applications. Hence, programs are needed that increase academicindustry collaboration for leveraging new devices (with industry provided software development kits or other interfaces), for enabling problem formulation and solving, and for designing more resilient infrastructure and services for applications envisioned for the year 2021.

Finding-6) Infrastructure that is broadly available should allow for testing of Year 2021 applications and services at scale. Driven by the data explosion as well as the consumerization of devices that brings more and more people online— not to mention devices/IoT— systems and applications need to be

designed to quickly expand and scale. Achieving this will require a focus on ease of use (simple configuration and setup), reproducibility of scientific computation and execution environments, local aggregation (clouds) with global interoperability and exchange, rethinking how researchers test systems (moving away from test==10 grad students), and a much shorter/faster research cycle (months, not years) that is based on interactive feedback from that testing.

- **Recommendation-1:** There is need for federated, deeply programmable, open, sliceable and sharable infrastructure for experimentation (similar to current GENI, CloudLab, Chameleon and beyond). Locavore infrastructure is especially needed at the edge for applications with lowlatency requirements. Additionally, the future experimental infrastructure should be designed to support broad categories of applications, not simply with particular "killer apps" in mind, as such predictions of the future tend to be incomplete. With broadly available infrastructure that can be operated as software-defined infrastructure, new solutions for national priority application areas can be realized. For e.g., they can be tightly integrated into smart manufacturing systems or dark factories. The vision would be for a software-defined dynamically reconfigurable factory with networked-robots making new products. Through careful measurement, the factory and its robots would proactively find/prevent faults and self-repair common problems to improve product quality and yield. This challenge would invoke big data analytics, control paradigms (cyberphysical systems), a reliable network of a very large number of sensors, and an architecture for a dynamically configured computing/networking infrastructure.
- **Recommendation-2:** Infrastructure should be designed and deployed in simple abstractions as well as with realistic application targets and use cases. Proposed approaches for infrastructure design should embrace the philosophy of minimal specificity (end-to-end), generic, and simple in order to maximize the generality of the resulting solution across multiple disciplines and user groups. Research and development of flexible and yet easy-to-use application programming interfaces that can handle heterogeneous devices and ownership are required. What are suitable abstractions and interfaces for devices, information streams, control loops, etc.? How do we create a set of components that can be shared and reused by many applications?

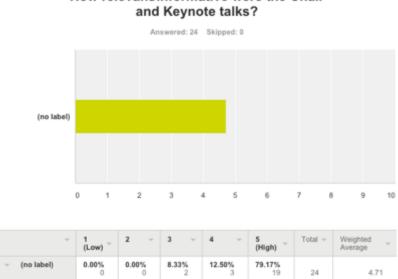
Finding-7) The infrastructure supporting applications and services in the year 2021 will take the form of an IoT that has unsolved security challenges due to: massive numbers of small devices, or devices with increasing data resolutions (relatively larger devices), each engineered to become inexpensive, massively distributed throughout physical space over time and more widely shared. Several factors such as: scale of devices, size of data sets, economy due to cheaper manufacturing methods, increased distribution and sharing—will all conspire to create new security challenges. In the IoT infrastructure for applications in 2021, the impact of security problems may be much worse than in IoC applications today. Increasing the numbers of devices disrupts the penetrate-and-patch paradigm the IoC depends on; connecting them intimately to physical infrastructure increases the physical consequences of compromise.

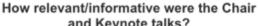
• **Recommendation-1:** <u>New techniques and related tools to make it easier to build distributed</u> <u>systems the right way quickly from a security standpoint should be developed.</u> The infrastructure of 2021 will require even more distributed systems, built even more quickly. Smart programmers today cannot build secure distributed systems quickly, and they need tools to make their development easier from a security standpoint. Also, they will frequently deal with dynamic systems whose interactions form control loops with each other and could lead to security loopholes and even system safety concerns. Moreover, the middleware infrastructure that is commonly used by application developers today is the web technology stack whose authentication/authorization assumptions do not scale well for services in a IoT world. Flexible and extensible web technology stack research needs to be supported to handle the technical. policy and legal specifications relating to infrastructure resources owned by multiple entities servicing users/devices at massive scales.

Recommendation-2: Investigation of innovative methods should be encouraged that allow automated security verification. Automation of security assurance is essential because it will be hard for human analysts and software engineers to answer: How can stakeholders (in reality, all users) have any kind of assurance that the software on these massive numbers of devices "works," from a security perspective? Are the interfaces free of standard lack-of-validation holes? Is use of the heap free of standard double-free and allocation overflow holes? Is the security policy sound---and fully mediated? Automation of security assurance tasks may allow scaling up in the speed and numbers of devices that can be verified. Successful solutions here will likely need to balance the advantage of full white-box testing with the privacy of source and design internals: successful solutions here will also likely need to figure out how to significantly increase assurance that this system-of-systems is secure, while recognizing the fundamental difficulty of "100%" confidence.

4. Post-Workshop Survey

There was a post-workshop survey that was sent to all the participants. About half of the participants provided responses. From the below survey responses, it is evident that the workshop was overall well received by the participants.



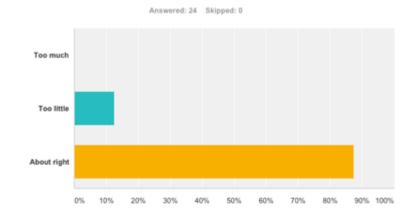


Did the Short Talks within the Breakout Sessions foster group discussions?



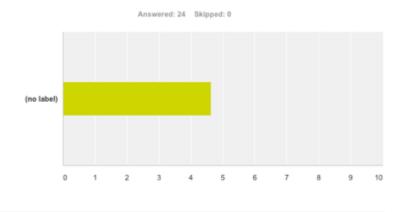
*	1 (Low) —	2 ~	3 ~	4 ~	5 (High) –	Total -	Weighted
 (no label) 	0.00% 0	4.17% 1	12.50% 3	37.50% 9	45.83% 11	24	4.25

Time allotted for presentations was...



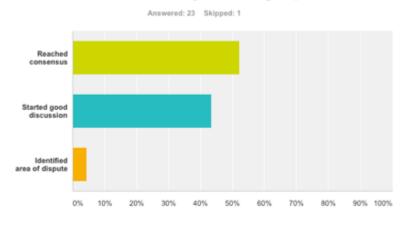
Answer Choices ~	Responses	~
- Too much	0.00%	0
- Too little	12.50%	3
 About right 	87.50%	21
Total		24

How relevant/informative were the group discussions within the Day-1 Breakout Sessions?



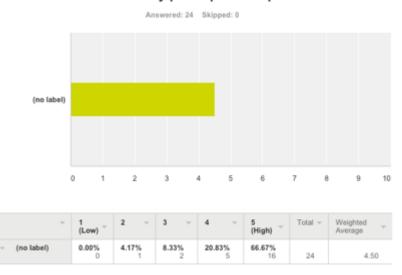
Ψ.	1 (Low) —	2	3 ~	4 ~	5 (High) —	Total -	Weighted
 (no label) 	0.00% 0	4.17% 1	4.17% 1	16.67% 4	75.00% 18	24	4.63

On most issues, my breakout group...

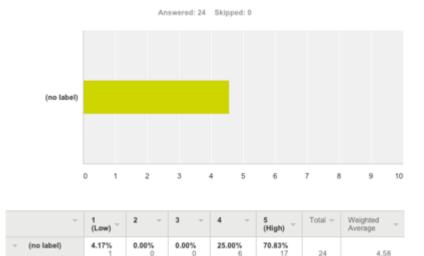


Ans	swer Choices v	Responses	*
Ŧ	Reached consensus	52.17%	12
Ŧ	Started good discussion	43.48%	10
Ŧ	Identified area of dispute	4.35%	1
Tot	al		23

Was the workshop format suitable for relevant/informative discussions and involved as many participants as possible?



Overall how satisfied were you with the workshop?



5. Conclusion

Applications and services based on computer science research are an integral part of the national innovation and economic development ecosystem that powers the United States economy and citizen quality of life. As just one example, the Internet alone has been adding 21% to economic growth over a recent 5-year period. However, by the year 2021, the architecture for today's web-oriented cloud-based Internet may well have reached full maturity and may no longer be contributing to growth at historical rates. Fortunately, we are now in the midst of a blossoming of new wireless and Internet technologies and growing human talent base; with *proper research funding and with attention to the applications and services that research can enable, there is a real opportunity to spur a whole new decade of improvement in our economy and quality of life.*

There were many interesting "imagine a future", "imagine a time/place" stories and innovative use of advanced technologies discussed at the workshop by invited speakers and between workshop participants in breakout discussion groups. In all these discussions, the community consensus and voice was clear -- that *NSF should continue to fund and consider expanding support for research that has broader impacts when applied to national interest applications and services*. In formulating research programs in CISE and other applicable NSF directorates, such broader-impact opportunities should be carefully considered. Challenge and grand challenge applications can often spur creativity and new insights for fundamental research. But, similarly, curiosity-driven research can be the catalyst for advances in economic competitiveness, healthcare, education and workforce development, clean energy, transportation, advanced manufacturing, and smart and connected communities.

Innovation in national interest applications and services can add value to fundamental and applied research in two very different ways: (i) they can translate the fundamental advances of curiosity-driven research into valuable advantages and benefits for Americans and the communities in which they live, and (ii) they can pose unsolved challenges that may suggest new and untraveled approaches for fundamental and applied research. Two examples can be considered in this context in terms of value creation. First is the NSF's GENI program investment that engendered and expanded the field of software-defined networking, which in turn has led to a vibrant industry sector in Internet and Cloud technologies in which American companies excel. Second is the research spawned by the NASA Apollo program, which posed unsolved "moonshot" challenges that ultimately ended up creating new materials such as Teflon and Velcro. Both modalities will help to add bold innovation and growth to the national economy and improve the quality of life for all Americans and for citizens in the rest of the world.

Appendix-I: Workshop Tracks Organization & Discussion Questions

Every breakout session should generate information on some general questions below as well as session-specific ones:

- o What are the latest advances and best practices? (Discussion Themes)
- o What are the open challenges? Which are missing definition? (Salient Findings)
- o What solution investigations are needed? What are higher priority? (Recommendations)

Track-1: Applications – Grand Challenges

(Scribe Lead: Nishal Mohan; Supporting Scribe: Glenn Ricart)

- What are some of the "imagine a future" use cases that future cyber infrastructure should support for national priority areas? You may consider some application areas such as e.g., Health Care, Education, Public Safety, Transportation, Earth Sciences, or Manufacturing.

- What are the top 5 application paradigms that may start seeing planetary-scale growth based on history and trends in related technologies convergence or changing social behaviors in the society? You may consider paradigms such as e.g., locavore clouds, self-federating systems, personalized cybercoaching.

- What are the factors that have kept some of the application areas (e.g., virtual reality) that for a long time showed promise but have not had larger impact? What challenges need to be addressed to foster scalable and sustained growth of these application areas in the 'year 2021 and beyond' evolution of the Internet?

- How can applications be instrumented with measurements and other data intelligence that enable them to better inter-operate, better predict and better adapt user QoE, especially when there are inter-dependent systems and processes?

- How will 'shared edge computing' or 'massive-scale public clouds' need to evolve for supporting applications and services in year 2021? You may consider evolution issues such as bit rates/latency, compute cycle access/pricing methods, service interfaces in terms of usability, or security tools for device-to-service or device-to-device communications.

Track-2: Data Intelligence and Visualization

(Scribe Lead: Micah Beck; Supporting Scribe: Prasad Calyam)

- What are the requirements of a next-generation data analysis and visualization systems to derive intelligence from e.g., "things", or social activity tracking of humans at a scale expected in year 2021?

- What are new challenges to be addressed (e.g., in discovery and search of data; verifying data veracity) to create actionable intelligence across multiple disciplines and user/provider communities? How can this intelligence be used beneficially to provide situational awareness or alter our environments in new ways?

- What are the gaps in approaches to co-operatively collect (e.g., for real-time analysis or historical data analysis) or compute (e.g., cloud/fog location selection) or share pertinent data sets (small, medium and big data) within user/resource federations that can be used for automated analysis, visualization and intelligence notification? You can imagine application use cases such as video content delivery, effective disaster response or cyber attack impact mitigation.

- What futuristic analytical and statistical techniques will be needed to deal with applications and services in the year 2021? You may consider topics in fields such as machine learning, optimization, computer vision, and expert systems.

Track-3: Services and Infrastructure

(Scribe Lead: Malathi Veeraraghavan; Supporting Scribe: Rick McGeer)

- What are the fundamental barriers and opportunities for new experimental methodologies that will shape the infrastructure and services for applications in the year 2021? You can consider issues that will influence applications in the year 2021 in the context of: next-generation emulators/simulators, increase of complex systems that are inter-dependent, ability to re-use proven configurations for repeatable science, or new ways of human-computer interactions.

- How will the convergence of growing advances in diverse domains (e.g., hardware at both edges and core, distributed computing frameworks, web technologies, autonomous management with software-defined infrastructure) enable new application realities in the year 2021? You may consider use cases, computing paradigms or technologies that will become possible that are in early stages today but have immense promise.

- What are the capabilities that need to be available to users, application developers and service providers that can improve ease and quantity/quality of "access" to infrastructure resources that are owned by multiple entities?

- What will be the nature of collaboration necessary between industry and academia to meet the infrastructure and service needs of applications in the year 2021? You may consider issues such as building blocks collaboratively developed within open software, open systems or other areas of open innovation.

Track-4: Security and Privacy

(Scribe Lead: Sean Smith; Supporting Scribe: Kaiqi Xiong)

- What are the major cyber security and privacy threats to be addressed for applications and services in year 2021? Will the concepts of security and privacy themselves change by year 2021? (How will the attack surface change?)

- How can we provide security and privacy across heterogeneous systems with varying levels of trust? You may consider cases where e.g., security and privacy alignment is needed across multiple organizations with their own respective resource and user policies?

- What are new security and privacy preserving and verification techniques that may be needed for applications and services in the year 2021? You may consider techniques e.g., artificial intelligence agents that can make or confirm information sharing, or techniques that are better than rule-based methods or future encryption approaches.

- What new transformative innovations and cultural changes in security and privacy are needed to deal with the coming "Internet of Things"? You may consider cases where humans cannot effectively reason about security and privacy when: devices become too long-lived, too cheap, too tightly tied to physical life in personal spaces, too invisible, and too many.

Appendix-II: Exemplar Application Grand Challenges

One of the best ways to motivate multi-party and inter-disciplinary collaborative research is to pose and solve Grand Challenge problems. A listing of Grand Challenges that were discussed in the Track-1 of the workshop are listed below. It is hoped that efforts to address the research and infrastructure recommendations presented earlier in this report will ultimately support effective and efficient solutions of such Grand Challenges. Additionally, the research and infrastructure efforts may require multi-agency and even international and/or industry collaborations in some cases.

Grand Challenge #1: Health Maintenance and Management

Improving the quality of life and decreasing healthcare costs are both possible with better health maintenance and management. CISE research can make a substantial difference by providing automated monitoring and measurement tools that can feed algorithms giving people advice on improving their health and calling for assistance when needed.

Having computers watch your behaviors seems to be an accepted practice. In addition to the growing use of fitness bands, cameras and sensors could aid with aging in place, providing more comfortable home environments for millions of seniors while making sure they have assistance when needed. Robots and other monitoring agents would feed algorithms that provide feedback to the senior and notify caregivers when outside assistance is needed. Seniors would want the comfort of knowing that people were not watching the cameras; only computers looking out for their welfare.

Chronic lifestyle diseases can be prevented in many cases. These include atherosclerosis, heart disease, and stroke; obesity and type 2 diabetes; and diseases associated with smoking and alcohol and drug abuse. Research could provide patient-specific real-time and longitudinal information needed to provide real-time behavioral feedback as well as longer-term cumulative reports for the person and their doctors and care-givers. Early indications are that using changes in easily-observable long-term continuously-recorded data functions well as indicators for chronic/acute disease propensity and onset. Because only computers are watching the data, there can be automated aggregation / summarization of data while maintaining patient privacy/security by using slices and private clouds.

Grand Challenge #2: <u>Realistic Virtual Reality (including augmented reality and mixed reality)</u>

Virtual, augmented, and mixed reality holds the hope of providing high impact educational and healthmaintenance experiences. Realistic, Holodeck-quality virtual and augmented reality triggers areas of the brain that reinforce learning. A small experiment with a learning game for 8th graders to teach them Cartesian Coordinates and slopes was far more effective and had longer-lasting impact than the same material taught in a traditional classroom and was far more engaging for the students. The possibility of using high-quality virtual reality for STEM education for higher-risk children and helping them into higher-paying higher-impact careers is compelling.

Accurate surgery visualization would be helpful for both surgeons practicing for surgery and for patients to understand the scope and risks of the surgery for which they are about to give informed consent.

An elderly population may appreciate the virtual mobility that realistic virtual reality could provide.

Cultures could be preserved across generations via accurate and detailed virtual reality.

Currently these kinds of applications can be demonstrated using expensive virtual reality headsets and high-powered graphics cards in computers tethered to the headsets. All of this creates an expensive solution. By using gigabit networking, VR could be streamed where needed for only the cost of the headset. Even that can be minimized by using a smart phone in a headset holder. Good results have been obtained with even Google Cardboard.

Grand Challenge #3: <u>Very secure, fault-tolerant critical applications infrastructure</u>

The banking and medical industries would greatly value secure, fault-tolerant critical applications infrastructure. How can and should such trustworthy systems be designed and constructed? Can software-defined slices provide additional security and privacy for sensitive information, perhaps in addition to encryption.

There would seem to be an opportunity for the financial industry to save billions lost to cyber-fraud if we could better protect sensitive streams, perhaps by putting them in dynamically-created and software-partitioned slices.

As the amount of cyberinfrastructure grows, could slicing and software-defined infrastructure provide a better building block than simply attaching more devices to a common Internet?

Could failures be handled as more routine items, becoming annoying retries instead of disastrous outages? Could the Netflix Chaos Monkey teach us more about how to create more fault-tolerant applications infrastructure?

Grand Challenge #4: Intelligent assistants and coaching

Better intelligent assistants and coaches could be significant productivity boosters for the economy and for helping people accomplish more than they could otherwise accomplish by themselves. Artificial intelligence now appears to be coming of age, and current results from Siri, Cortana, and Google Now show that there are tremendous opportunities to do such things as:

- a. create records of creative work
- b. capture information for automatic collaboration and sharing
- c. assist with maintaining security and privacy
- d. help fight negative addictive behaviors
- e. provide assistance with just-in-time learning
- f. assist with giving individual advice (such as how to escape) during disasters or other emergencies
- g. turning downtime hours into more enjoyable "quality time"

Grand Challenge #5: Erase distance barriers

Virtual collaboration capabilities also appear to be coming of age. We need research on how they could be more naturally used without experts, be more reliable, have better identity management, include features for privacy (e.g., for medical consultations), have less latency and annoying pauses, could take better advantage of unconstrained bandwidth, and allowing recipients to choose their attention points and points of view (such as selecting individual speakers in a conference when multiple people are talking).

Many services could be better delivered to rural and first nation areas by "streaming" medical / educational capabilities.

Grand Challenge #6: Two "Covering" Challenges: Disaster Management and Urban Sciences

Two "covering" challenges can motivate requirements in many areas and could be viewed as an "engage everyone" type of challenge. The first in this category is disaster preparedness and response management. It can engage multiple disciplines and multiple areas of computer science in a single grand challenge. However, we should not only look at disastrous scenarios, but also look at more mild or "everyday" disasters as well since they occur far more frequently. The second is the broad area of Urban Sciences. It can also engage multiple disciplines and multiple areas of computer science in a single but encompassing subject. Urban sciences could include cyberphysical systems, people-oriented systems, the interaction of people with their physical and digital environments, and systems to help improve quality of life. Our communities would serve well as living laboratories to understand some of society's most pressing issues and suggest both preventative and remedial intervention.