Cybersecurity Experimentation of the Future (CEF): Catalyzing a New Generation of Experimental Cybersecurity Research

Community Plan and Roadmap to Develop Future Experimentation Infrastructure in Support of Cybersecurity Research

FINAL REPORT
July 31, 2015

Cybersecurity Experimentation of the Future (CEF): Catalyzing a New Generation of Experimental Cybersecurity Research

Community Plan and Roadmap to Develop Future Experimentation Infrastructure in Support of Cybersecurity Research

FINAL REPORT

July 31, 2015

David Balenson and Laura Tinnel, SRI International
Terry Benzel, USC Information Sciences Institute

SRI International®

USC Viterbi School of Engineering
Information Sciences Institute

This material is based upon work supported by the National Science Foundation under Grant No. ACI-1346277 and ACI-1346285. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

EXECUTIVE SUMMARY

This report presents a strategic plan and enabling roadmap intended to catalyze generational advances in the field of experimental cybersecurity research. These results represent the conclusions of a study conducted under NSF auspices by SRI International and USC Information Sciences Institute throughout calendar year 2014. The study had broad participation by stakeholders representing the cybersecurity research, research sponsor, and customer communities. The report outlines the process and methodology of the project, presents key inputs, supporting evidence developed through the course of the study, and synthesized results, and then presents our final conclusions.

Our overarching finding is that transformational progress in three distinct, yet synergistic, areas is required to achieve the desired objectives:

1) Fundamental and broad intellectual advances in the field of experimental methodologies and techniques, with particular focus on complex systems and human-computer interactions.
2) New approaches to rapid and effective sharing of data and knowledge and information synthesis that accelerate multi-discipline and cross-organizational knowledge generation and community building.
3) Advanced, accessible experimentation infrastructure capabilities.

The central result of our study is a roadmap that presents requirements, objectives and goals in each of the areas outlined above over three, five and ten year phases. In some cases, the phases build upon each other, and in other cases, new fundamental research is required over a longer period of time to satisfy the objectives of the roadmap.

Taken together, these areas, as embodied in the roadmap, paint a vision for a new generation of experimental cybersecurity research – one that offers powerful assistance towards helping researchers shift the asymmetric cyberspace context to one of greater planning, preparedness, and higher assurance fielded solutions.

The capabilities identified in the roadmap take into account the current state of the art in experimental cybersecurity research and its supporting infrastructure, other types of research facilities, and existing cyber-domain "test and evaluation" capabilities. In addition to leveraging current and expected capabilities in cybersecurity and adjacent areas, the roadmap presumes advances in key computer science disciplines such as ontologies, metadata, libraries, and corresponding resource discovery.

We emphasize that while this type of study would typically focus heavily on experimentation infrastructure (i.e., tools and testbeds), and while we did pay significant attention to this topic, our fundamental conclusion is that an emphasis on infrastructure alone will fall far short of achieving the transformational shift in the research, community, and experimentation required to address cybersecurity in the rapidly changing cyber environment.

Our conclusion is that strong, coupled, and synergistic advances across each of the areas outlined above – fundamental methodological development, fostering and leveraging communities of researchers, and in the capabilities of the infrastructure supporting that research – will transform the field of cybersecurity.

TABLE OF CONTENTS

1 INTRODUCTION

This report presents a strategic plan and enabling roadmap intended to catalyze generational advances in the field of experimental cybersecurity research.

These results represent the conclusions of a study conducted under NSF auspices by SRI International and USC Information Sciences Institute (USC-ISI) throughout calendar year 2014. The study had broad participation by stakeholders representing the cybersecurity research, research sponsor, and customer communities. The report outlines the process and methodology of the project, presents key inputs, supporting evidence developed through the course of the study, and synthesized results, and then presents our final conclusions.

Our overarching finding is that transformational progress in three distinct, yet synergistic, areas is required to achieve the desired objectives:

1) Fundamental and broad intellectual advances in the field of experimental methodologies and techniques, with particular focus on complex systems and human-computer interactions.
2) New approaches to rapid and effective sharing of data and knowledge and information synthesis that accelerate multi-discipline and cross-organizational knowledge generation and community building.
3) Advanced, accessible experimentation infrastructure capabilities.

The central result of our study is a roadmap that presents requirements, objectives and goals in each of the areas outlined above over identified three, five and ten year phases. In some cases, the phases build upon each other, and in other cases, new fundamental research is required over a longer period of time to satisfy the objectives of the roadmap.

Taken together, these areas, as embodied in the roadmap, paint a vision for a new generation of experimental cybersecurity research – one that offers powerful assistance towards helping researchers shift the asymmetric cyberspace context to one of greater planning, preparedness, and higher assurance fielded solutions.

The capabilities identified in the roadmap take into account the current state of the art in experimental cybersecurity research and its supporting infrastructure, other types of research facilities, and existing cyber-domain "test and evaluation" capabilities. In addition to leveraging current and expected capabilities in cybersecurity and adjacent areas, the roadmap presumes advances in key computer science disciplines such as ontologies, metadata, libraries, and resource discovery.

We emphasize that while this type of study would typically focus heavily on experimentation infrastructure (i.e., tools and testbeds) and while we did pay significant attention to this topic, our fundamental conclusion is that an emphasis on infrastructure alone will fall far short of achieving the transformational shift in the research, community, and experimentation required to address cybersecurity in the rapidly changing cyber environment.

The study results point to a new direction for the field of experimental cybersecurity research and development. The importance of research into *the science of cybersecurity experimentation* is an overarching need. Any set of requirements or capabilities for cybersecurity experimentation must be backed by transformational progress in the science

of experimentation. It is only by grounding our research in scientific methods and tools that we can realize the impact that experimentation can have. It should be noted that this call for research into the science of cybersecurity experimentation is different from the current fundamental research into the science of cybersecurity, though they are certainly complementary in their eventual goals. Along with establishing a field of research into the science of cybersecurity experimentation, substantial new approaches to sharing are needed in order to enable scalable, cross-discipline experiments. Needed new approaches to sharing include all aspects of the experimental science, from data, to designs, to experiments to the research infrastructure itself. Finally, in order for this new field to take shape and have significant impact, a cultural shift in the way researchers approach experimentation is required. Researchers must move towards defining and running repeatable, hypothesis-based experiments using community-driven experimentation infrastructure and components.

Our conclusion is that strong, coupled, and synergistic advances across each of the areas outlined above – fundamental methodological development, fostering and leveraging communities of researchers, and in the capabilities of the infrastructure supporting that research – will transform the field of cybersecurity.

1.1 ROADMAP FINDINGS

Cybersecurity challenges today are very real and wide-ranging with significant implications across most critical sectors of our society. In order to address this situation cybersecurity experimentation needs to be applicable across multiple domains and communities. Cybersecurity is no longer the sole purview of computer scientists and engineers; it needs to also be accessible to researchers in critical infrastructure domains such as energy, transportation, manufacturing, finance, healthcare, economics, human behavior, and many others. The details of this recommendation are described in the roadmap Section 4.1, Domains of Applicability.

For research in cybersecurity to be impactful, it must be based on both sound science and on the real world. The community needs shared, validated models and tools that help researchers rapidly design meaningful experiments and test environments. These models are needed for both real and simulated test environments. The requirements in this area are described in the roadmap Section 4.2, Modeling the Real World for Scientifically Sound Experiments. The need for experimentation grounded in the real world is also discussed below as one of the foundational principles for the roadmap.

Architectures are needed that provide generic experimental research frameworks that allow for specialized, domain-specific instantiations. This architectural principle is more important than the actual connection fabric and it moves the discussion of research infrastructure from one of "realization" mechanisms to the higher-level design and representation requirements for experimental research.

Architectures are also needed that will enable multi-discipline experimental research in cybersecurity. This leads to fundamental questions in experiment design and scenario exploration. In addition, there are a number of important engineering challenges in connection fabrics, including resource discovery and instantiation, These topics are discussed in roadmap Section 4.3, Frameworks and Building Blocks for Extensibility,

Section 4.4, Experiment Design and Instantiation, and Section 4.5, Interconnected Research Infrastructure.

Research in cybersecurity requires sound science that builds upon controlled, well-executed experiments. The community needs tools that will help researchers run and control experiments in interactive test environments that are fully orchestrated, provide customized visualizations and interactions, and support debugging and validation. These requirements are described in roadmap Section 4.6, Experiment Execution and Management.

Research in cybersecurity requires the collection, ensured integrity, and analysis of experimental data. Instrumentation and data collectors, transport, and protection mechanisms should be non-intrusive. Also needed are data repositories and reusable analysis tools and techniques. These requirements are described in roadmap Section 4.7, Instrumentation and Experiment Analysis.

In addition to the core capabilities described in the roadmap, there are several important meta-properties that underlie all other needed capabilities. Research infrastructure needs to be easily usable by both a wide range of experimenters and by the owner/operators of the infrastructure. Then, in support of experiment validity and as described in roadmap Section 4.6, the community needs mechanisms and processes to provide confidentiality, availability and integrity of the experiment ecosystem. Finally, there are a number of cultural and social changes, along with community building, that are needed to facilitate future capabilities. These requirements are described in roadmap Section 4.8, Meta-Properties.

1.2 TOP FIVE RECOMMENDATIONS

An analysis of the complete roadmap resulted in a set of Top Five Recommendations. We believe that active engagement in these five areas identified in the roadmap will yield impactful transformational results; it will form the needed basis for further research, development and integration of the remaining roadmap areas and new emerging topics not yet captured. It is important to note that while the top five are most deserving of immediate attention, all of the roadmap areas are vitally important to achieving the roadmap findings.

Recommendation 1: Domains of Applicability – Multidisciplinary Experimentation

The creation of specialized experimentation capabilities across multiple domains is already a fast growing area of investment and research. In the near term, adding a focus on multidisciplinary experimentation that includes computer science, engineering, mathematics, modeling, human behavior, sociology, economics, and education will have the greatest impact on accelerating cyber security experimentation. These new capabilities will allow researchers to address open areas around the pervasive nature of cyber security and can provide an avenue to address emerging issues surrounding the Internet of Things (IoT). A focus in this area, along with the other top recommendations, will create the opportunity to coalesce the variety of emerging advances in capturing the human element in other domains. The details supporting this recommendation are in Section 4.1.

Recommendation 2: Modeling the Real World for Scientifically Sound Experiments – Human Activity

The ability to accurately represent fully reactionary complex human and group activity in experiments will be instrumental in creating laboratory environments that realistically represent real-world cyber operations. To date most cyber experimentation is conducted in closed environments with minimal synthetic representation of human behavior. In order for this area to provide transformational results it must include the ability to integrate live and synthetic humans without artificialities that may interfere in some experiments, as well as capabilities to help ensure scientific validity when including live humans in experiments. Introducing the human element in experimentation will also open up the door to privacy and ethics issues that must be addressed. Achieving a seamless blending of the cyber and human world experimentation is a high priority mid term activity. The details supporting this recommendation are in Section 4.2.

Recommendation 3: Frameworks and Building Blocks for Extensibility – Open Interfaces

Creating open standards and interfaces, for both experimental infrastructure facilities and for experiments themselves, is a high priority mid-term activity. Developing common models of infrastructure and experiment components to open interfaces and standards contributes to the overall goal of fostering a field of the science of cybersecurity experimentation. As a result, communities will be able to conduct, validate, integrate and share experiments, experimental components, and experimental results. This fundamental ability is needed to enable broader research in cybersecurity, as opposed to working in narrow sub disciplines. In addition, this new sharing capability will enable researchers to more easily repeat peer experiments and build upon those results. The details supporting this recommendation are in Section 4.3.

Recommendation 4: Experiment Design and Instantiation - Reusable Designs for Science-based Hypothesis Testing

Research, development, and exploration in the area of experiment designs and design patterns for science-based hypothesis testing are required in order to achieve transformational changes in the field of experimental methodologies and techniques for cybersecurity. Researchers across all domains that rely on computational systems must be able to rapidly design meaningful experiments that reflect the real world by reusing and extending existing, validated experiment designs and design components. Experiment designs should be automatically validated and processed akin to the use of software development environments. Advances in key computer science disciplines such as ontologies, meta-data, libraries, and resource discovery are necessary to realize highly automated, extensible, and validated experiment designs. The details supporting this recommendation are in Section 4.4.

Recommendation 5: Meta-properties – Usability and Cultural Changes

Cybersecurity research infrastructure must be usable by a wide range of researchers and experts across many different domains of research and not limited to traditional computer science researchers. It is vital that experimental capabilities not be restricted to power users of cybersecurity experimentation infrastructure. Given that the future research

infrastructure is envisioned to be dynamic and used in many different ways, it is important to consider not only the usability of the technologies used to administer infrastructure, but also that of the technologies used to create and deploy experimentation infrastructure. In the long term we envision some degree of self-configuring or infrastructure-in-a-box capabilities to ease operational burdens, particularly for prospective researchers who are not from a traditional computer science background.

Along with usability-related properties, the adoption and use of future experimentation infrastructure likely will be characterized by the use of evolvable frameworks that support advances in experimental methods, and multiple models of collaboration, within both the user base of a single infrastructure and collaboration that spans multiple infrastructures. Research activities in several core capability areas discussed in the roadmap are required. Researchers must make a concerted effort to take advantage of community based resources, rather than relying on homegrown approaches. The shift to multi-domain users and the use of shared frameworks will enable both the research infrastructure and cybersecurity researcher communities to co-evolve. Usability and cultural changes are a long-term priority on which work must begin immediately. The details supporting this recommendation are in Section 4.8.

1.3 DEFINITION OF "CYBERSECURITY EXPERIMENTATION INFRASTRUCTURE"

The roadmap prescribed by this effort is based on a definition of *cybersecurity experimentation infrastructure* that encompasses general-purpose ranges and testbeds that are both physical and/or virtual in nature. These testbeds may be for generalized use or may possibly be highly specialized, such as certain cyber physical testbeds, and may include specific physical apparatus, hardware tools, and simulators.

These test facilities are more than just physical and/or virtual systems; they include software tools that support one or more parts of the experiment life cycle, including, but not limited to:

- Experiment design
- Testbed provisioning software
- Experiment control software
- Testbed validation
- Human and system activity emulators
- Instrumentation – systems and humans
- Data analysis
- Testbed health and situational awareness
- Experiment situational awareness
- Experiment designs, models, and results
- Other similarly relevant tools

Many of the above items are multi-purpose, meaning they are applicable not only in experimental research, but also in conducting Test and Evaluation (T&E), Independent Verification and Validation (IV&V), operational training exercises, and educational activities. As capabilities in these tools mature, they will be useful to also satisfy the requirements for T&E, IV&V, training, and education. Conversely, existing tools used to support T&E, IV&V, training, and education may be applicable in cybersecurity

experimentation; where such is the case, these tools can be adopted for used in experimentation.

However, research-oriented experimentation requires capabilities that extend beyond what is needed to support these other activities. Given the central premise of fundamental methodological development, fostering and leveraging communities of researchers, and revolutionary advances in the capabilities of the infrastructure supporting that research, it is important to understand that research infrastructure is by its very nature more than the infrastructure of machines and tools. Research infrastructure encompasses scientific methodologies, experimental processes, and education that are critical to effective use of machines and tools. Specifically, research infrastructure requires meta-research into:

- Design specifications (multi-layered languages and visualization)
- Abstractions and effective abstraction methodologies and techniques
- Approaches to the semantic analysis and understanding of experimenter intent
- Constraints and formal methods for determining experiment validity

Lastly, we acknowledge the critical community need for good data sets in cybersecurity experimentation, such as those provided by the Protected Repository for Defense of Infrastructure Against Cyber Threats (PREDICT) [1], however, the focus of this effort is on the tools and methodologies needed to design, execute, analyze, and share scientifically rigorous experiments in support of cybersecurity research.

1.4 WHERE IS EXPERIMENTATION APPLICABLE?

Understanding both the motivation and the areas in which experimentation is applicable is important to developing requirements for experimental infrastructure in the future.

If experimental cybersecurity research is to have a role in shifting the asymmetric cyberspace environment, then its supporting infrastructure must help increase researcher effectiveness and support the generation and preservation of solid empirical evidence. Research infrastructure should be designed to enable research, not constrain it. All too often researchers are required to work in experiment facilities that, due to limitations of the available test apparatus, place limits on the research question that can be explored.

Further, new mechanisms and processes are needed to capture and share knowledge. Sharing of experimental designs, data, and results serves two key purposes: 1) it enables peer review and repeatability, thereby increasing the scientific soundness of research conducted using the infrastructure, and 2) it enables researchers to build on the sound results of other researchers, thereby increasing the number of solutions available to address hard cybersecurity problems.

For purposes of this study and roadmap, it is assumed that experimentation is about learning and is used for a wide variety of investigations:

- To perform an evaluation
- To explore a hypothesis
- To characterize complex behavior
- To complement a theory
- To understand a threat
- To probe and understand a technology

1.5 REPRESENTATIVE CYBERSECURITY HARD PROBLEMS

As a first step in examining requirements for future experimental infrastructure, we identified classes of representative hard problems in cybersecurity that would be amenable to experimental research. Cybersecurity (and hence cybersecurity experimentation) is intrinsically hard. It often involves analysis of large, complex, decentralized systems. Furthermore, cybersecurity focuses on worst case behaviors and rare events, often in the context of multi-party and adversarial/competitive scenarios.

Experiments and scenarios must be sufficiently well framed, scaled and realistic (see Real World discussion below for more on realism) to be valid. When they are not, they are instead misleading. It is therefore not sufficient to enumerate a set of "hard problems" for cybersecurity experimentation, but one must also consider the above facets to ensure that the problems are amenable to meaningful experimentation.

The identified hard problems were drawn from several research communities and, not surprisingly, tended to reflect current topic of interest. Nonetheless, they provide a sounding board for which we may project needed experimental infrastructure. Working with researchers in the community, we identified the following broad categories of hard problems in the areas of systems and software, networking, and cyber-physical systems.

In the area of *systems and software*, future challenges that are considered to be amenable to experimentation, include:

- Human Interactions
- System-of-systems security metrics
- Emergent behavior in large scale systems
- Supply chain and roots of trust
- Societal impacts and regulatory policies

In the area of *networks*, future challenges that are considered to be amenable to experimentation, include:

- Anonymity and privacy of data and communication
- Trust infrastructure
- Software defined networking and other unbundling of network function
- Political, social, and economic (balance-of-interest) goals in network design
- Pervasive communications, across multiple organizational, political boundaries

Cyber-physical systems are an emerging area of interest for cybersecurity and experimental research particularly in the areas of converged cyber and physical phenomena. Future challenges that are considered to be amenable to experimentation, include:

- Embedded devices
- Autonomous vehicles, smart transportation
- Electric power, smart grid
- Medical implants, body sensors, etc.

1.6 EXPERIMENTATION – IT'S ABOUT THE REAL WORLD

For experiments to be relevant, they must start with the models of the real world. This introduces a fundamental set of research topics in modeling and abstraction. All experimentation starts with some form of conceptual model that is then captured in an experiment with varying degrees of abstraction and validity. As a result there are a number of key research areas ranging from experiment design specifications to auto-generated model refinement, and a range of methodologies and tools to assess representation.

The concept of "realistic" may well be the most poorly defined concept in the whole systems testbed and experimentation infrastructure domain. When an experimenter claims that her environment needs to be realistic, the next question is, "*In what axes, and what non-axes, must it be realistic? Models of users? User behaviors? CPU performance? Economics? Industry structure? Operations and management skills? Other?*" Additional open questions further complicate the issue of realism: "*How does one quantify/evaluate "realistic enough"? How does one avoid studying the problems she already knows?*"

Thus, to drive future experimentation in the realism dimension, we require challenge problems that illustrate and motivate taxonomies of realism and offer insight into metrics for realism sufficiency. This need relates back to the discussion of hard problems above and is vital to any discussion of real world experimentation through modeling. Assuming that future experimental infrastructure includes support for modeling real world systems, then additional methods and tools can help extend these modeling activities to provide increasing forms of real world models.

Industry and the private sector can play a key role in satisfying realism requirements. They have the direct experience with key infrastructure, data, and components that are either the subject of an experiment or serve as part of the experimentation infrastructure itself. The community should reach out to industry, wherever possible, and find ways to collaborate and benefit from industry contributions of models, specifications, prototypes, data, and, where appropriate, components (e.g., routers). As beneficial as industry involvement can be, it is important to not become too tied to the specifics of today's systems; the cyber world evolves rapidly, so solutions based on today's systems may have little applicability five years from now. Experimental research must be about the future and how things might be, so we must keep our eye on projections for the future.

1.7 MOTIVATION: WHY ARE WE DOING THIS?

Society's cyber dependencies are rapidly evolving. In nearly every aspect of our lives, we are moving toward pervasive embedded computing with a fundamental shift in network properties. These changes bring a very real and wide-ranging set of challenging cyber threats. Addressing these challenges will require cybersecurity research based on sound scientific principles. The scale and complexity of the challenges will require that researchers apply new experimentation methods that enable discovery, validation, and ongoing analysis.

Cybersecurity research and development is still a relatively young field and as discussed above, poses certain intrinsically hard challenges due to the inherent focus on worst case behaviors and rare events, in the context of multi-party and adversarial/competitive scenarios. Research infrastructure is a crucial piece of the puzzle, providing environments

where new cybersecurity research hypotheses can be tested, stressed, observed, reformulated, and ultimately proven before making their way into operational systems.

The ever increasing cyber threat landscape demands new forms of advanced research and development and in parallel new revolutionary approaches to experimentation and test. While the current state of the art in cybersecurity experimentation has recently had increased focus and investment, there is clearly a need for future research infrastructure that can play a transformative role for cybersecurity research well into the next decade.

1.8 AUDIENCE OF THE REPORT

The intended audience of this report is both the cybersecurity research community and the research infrastructure community. Bridging these two communities through a common view of experimentation needs and the research and development necessary to achieve that vision is vital to making true advances in the security of our cyber systems. Intended audience members are expected to come from academia, private industry, government funding agencies, and leading cyber and enterprise organizations.

As a part of the NSF community aspect of this project, a number of study groups were held that contributed to the formation of the concepts and specifics of this report. This draft report will be widely circulated amongst study group participants, industry leaders, government organizations and a wide range of researchers and developers.

1.9 STRUCTURE OF THE REPORT

The remainder of this report is structured as follows. Section 1 describes the study groups, their agenda, participants and output. Section 1 describes the survey of existing cybersecurity research infrastructure. This survey examined a wide range of testbeds and tools and developed recommendations for where current infrastructure might be leveraged to address requirements in the roadmap. It identified areas where significant new work is needed. Section 4 presents the roadmap outlining requirements and goals for future research infrastructure over three, five and ten year phases. The roadmap was developed as a synthesis across study group input, assessment of the state of the art, and principal investigator experience. Section 1 presents conclusions and summarizes community recommendations. Finally, we end with acknowledgements, references, and appendices with details of the existing infrastructure survey, study groups, and study group participants.

2 STUDY DESCRIPTION

2.1 OVERALL PROCESS AND APPROACH

Members of SRI International's Computer Science Laboratory (SRI) and the University of Southern California's Information Sciences Institute (USC-ISI) conducted the CEF study as a collaborative effort, with broad participation by members of the cybersecurity research, research sponsor, and customer communities. Figure 1 depicts the overall structure and organization of the study effort.

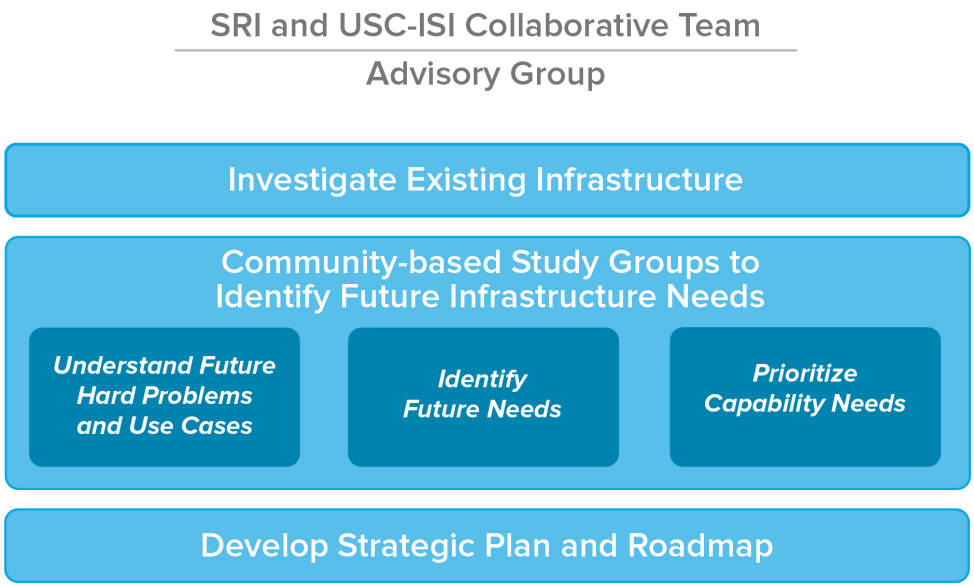


Figure 1. Community-based effort to develop a strategic plan and roadmap to guide the development of experimentation infrastructure to enable research into future cybersecurity challenges.

An Advisory Group (AG), comprised of seven senior leaders from government, industry, and academia, helped inform and guide our work (see Appendix C). The AG helped ensure the study focused on important, forward-seeking questions and issues, and that it engaged a wide set of participants from relevant areas and disciplines.

The study included three main thrusts. In the first thrust, we investigated existing experimentation infrastructure and user community experiences to gain an understanding of current capabilities, potential limitations, and future directions.

During the second thrust, we explored future cybersecurity experimentation infrastructure needs through a series of three community-based study groups. The study groups were explicitly organized to: 1) understand hard cybersecurity problems and use cases that could benefit from experiment-driven research, 2) identify the experimentation infrastructure needed to facilitate research focused on addressing the hard problems and use cases, and (c) identify gaps between needed and current capabilities and prioritize capabilities based on domain needs.

The third thrust synthesized requirements and needed capabilities into the strategic plan and enabling roadmap. The synthesis gave rise to a set of overarching findings. It became clear at the outset of the study that research infrastructure encompasses far more than just test apparatuses. Thus, the synthesized requirements point towards a roadmap that includes fundamental advances in the field of experimental methodologies, new approaches to accelerating multi-discipline and cross-organizational community-based experimentation, and the development of a new generation of sustainable experimentation infrastructure. The resulting infrastructure is intended to support the revolutionary advances in the field of experimental cybersecurity research discussed earlier.

2.2 INVESTIGATE EXISTING EXPERIMENTATION INFRASTRUCTURE

The knowledge of existing experimentation infrastructure, facilities, and user experiences is key to understanding the capability gaps that inform future cybersecurity experimentation. The first phase of this study included an investigation of the capabilities, potential limitations, and plans for future enhancement of existing experimentation facilities, including DETER [2], GENI [3], and National Cyber Range (NCR) [4][5]. This phase of study was not intended to provide a comprehensive deep survey but instead consisted primarily of a study and analysis of on line resources, previous studies, and web searches. The advisory group provided some leads and suggestions for additional areas of investigation. This phase of the study is described in Section 1. The results of the investigation into existing research infrastructure served as input for our capability needs study group.

2.3 CONDUCT COMMUNITY-BASED STUDY GROUPS

The CEF team led three study groups involving the broad community as stakeholders. Each of the study groups built on the previous one to some degree, and there was a small amount of overlap between attendees as well as the organizers to help carry forward the results between the groups. The goal of the first study group was to understand hard cybersecurity problems and develop experimentation use cases. The participants were primarily researchers in cybersecurity who might be candidates for using future experimentation infrastructure. The study group consisted of several plenary talks and panels and then was broken into three working subgroups focused on systems/software, networks, and cyber-physical systems. The groups were asked to identify cybersecurity hard problems and future research that would be amenable to or benefit from experimentation and to define use cases around approaches to cybersecurity experimentation of the future. This group was asked to examine these questions independent of any particular research infrastructure.

The second study group was tasked with identifying future experimentation infrastructure requirements. The results of the first study group were presented along with several panels intended to stimulate the discussion. We continued the sub group topic areas of systems/software, networks, and cyber-physical systems from the first study group. The break out groups for Study Group 2 were asked to define approaches to cybersecurity experimentation of the future and to identify infrastructure and services needed to support the experiment types and objectives as defined.

The third study group was tasked with defining prioritized domain-based capability needs. The results of the first and second study groups were presented along with several panels

intended to stimulate the discussion. The study group was broken into three working groups focused on the topics of experiment design, experiment realization, and experiment analysis. The break out groups were asked to identify infrastructure and services needed to support the experiment types and objectives expected in the future, and to identify and prioritize missing cybersecurity experimentation capabilities needed over the coming years, for research challenges spanning multiple sectors.

The study group agendas and participants are included in Appendix B.

2.4 GENERATE STRATEGIC PLAN AND ROADMAP

The final phase of the study was the generation of a strategic plan and roadmap. We held several brainstorming sessions with an expanded team from both organizations. We also presented the work to several strategic stakeholders – National Science Foundation Directorate for Computer and Information Science and Engineering, Department of Homeland Security Science and Technology Directorate S&T, Army Research Laboratory, Cyber Security and Information Assurance Interagency Working Group [6], IFIP WG10.4 on Dependable Computing and Fault Tolerance [7], and Institute for Information Infrastructure Protection [8]. This phase consisted of synthesis of requirements and needed capabilities. The synthesis gave rise to a set of over arching findings and requirements that point towards a roadmap inclusive of fundamental advances in the field of experimental methodologies, new approaches to accelerating multi-discipline and cross-organizational community building and finally the development of a new generation of sustainable experimentation infrastructure.

The goal of the roadmap is not to prescribe the creation of a single instance of cyber experimentation facility or a static definition of a testbed. Rather, taken together, these paint a vision for a new generation of experimental cybersecurity research that offers powerful assistance towards shifting the asymmetric cyberspace context to one of greater planning, preparedness and higher assurance fielded solution.

3 SURVEY OF EXISTING INFRASTRUCTURE

The CEF team conducted a survey of existing research infrastructure to understand the current state of the art in technologies used in testbeds and laboratories within the United States. We summarize our approach and findings here.

3.1 APPROACH

We surveyed mostly in cybersecurity research infrastructure in the United States, although we did consider two sets of infrastructure outside the United States. Initially we identified 46 candidates for the final survey. For each of these candidates, we used Internet searches to obtain a preliminary set of information that was then used to narrow down the survey set to a final, representative set of the infrastructure that is either commonly available or could provide significant value if made available.

The final set of surveyed infrastructure includes generic testbeds with open source tools, special purpose testbeds with proprietary tools, and specific tools that are not part of any particular testbed. The final survey set was divided amongst a team of researchers with test and experimentation experience at both USC-ISI and SRI. Each team member interviewed one or more key people at the owning organization and filled out a survey template. In some cases, in person visits were made to see the infrastructure in operation in its native environment. The completed survey templates may be found in Appendix A.

3.2 EXISTING TESTBEDS

Our survey found that a number of testbeds presently exist, however with a few exceptions, they are mostly proprietary and/or closed. The vast majority of testbeds we surveyed were focused on traditional information technology (IT) systems and networks, which does not address our nation's growing dependency on specialized embedded systems. Several of these IT focused testbeds had advanced capabilities but were military purposed and unavailable to support research by industry and academia. Other IT focused testbeds (e.g., DETERLab [9] and ProtoGENI [10]) are available for use by industry and academia. The Open Networking Lab (ON.Lab) [11] is a specialized testbed for software defined networking (SDN); it is available exclusively to members of the ON.Lab community.

Some new specialized testbeds in domains such as electric power and transportation exist, however they are exceptionally few in number. National labs and university consortiums typically operate SCADA or ICS testbeds and do not make them available for general research use. The Department of Transportation (DoT) operates a set of testbeds [12] comprised of physical roads to support research in developing a smart transportation system , and the University of Michigan recently announced their Mcity environment for testing driverless cars [13]. While not focused specifically on the cybersecurity issues in a smart transportation system, these testbeds may be useful for such purpose.

3.3 EXISTING TOOLS

While a fair number of "one off", proprietary research infrastructure tools and tool supporting capabilities currently exist, only a few are available for general use at either low or no cost. Most of these tools fall in the following categories:

- Testbed provisioning and control
- Standard IT network topology design
- Virtualization
- Experiment design specification languages

Other useful experimentation capabilities exist in government-funded national labs and within military-supporting test and evaluation facilities (e.g., Test Resource Management Center National Cyber Range (NCR) [4][5], Air Force Research Laboratory Cyber Experimentation Environment [14], Department of Energy National Laboratory testbeds) but are not generally available for industry and academic research. Such capabilities include advanced traffic generation tools, testbed network and system instantiation tools with self-validation, testbed situational awareness tools, and data collectors. Some of these tools were developed in closed, classified spaces and cannot be extracted for use elsewhere. Others were developed in unclassified spaces using government funding and are available for government contractor use only. Some tools, such as those developed at MIT Lincoln Laboratory [15], are currently under consideration for release to support non-government sponsored research.

It should be noted that a small number of commercial traffic generators do exist; however, they typically are cost prohibitive for use in academic research. Finally, the U.S. Government is presently sponsoring research in the following areas:

- Ontologies for rich experiment specification
- Graphical user interface based experiment design tools
- Advanced human emulation for background activity

Should this research prove fruitful, new capabilities in these areas may be available in the very near future.

3.4 SUMMARY

Based on the results of our survey, we conclude that a large amount of the experimentation infrastructure and capabilities needed to support cybersecurity research either does not exist or is not generally accessible to the broader research community. Restricted government use tools, such as those in the NCR, could provide additional coverage of the space of needed capabilities. Should the government choose to make these tools available for general research use, the needed investment in research infrastructure would be lower.

As a result of the lack of available infrastructure, a large portion of cybersecurity research efforts require researchers to design and create a full test apparatus from scratch. The process to build such is laborious, consuming valuable research dollars and time. Worse, these test apparatuses are seldom shared or reused by other researchers. While we do not believe that a set of shared community research infrastructure will completely obviate this practice, we do believe it will provide a basis upon which to build, effectively reducing the overall time and money spent on this portion of research efforts.

Finally, by creating one-off test apparatuses, we introduce the issue of testbed software correctness as an uncontrolled variable in experimentation. In doing so, we significantly increase the probability of errors in experimentation. Test apparatus errors can invalid results, wasting research dollars. Worse, if erroneous results are published, an error's effect could amplify when other research efforts are based on the erroneous work. Ultimately, invalid experimental results that are not caught can find their ways to vendors and result in weak or ineffective cybersecurity products. A shared, vetted community experimentation capability can reduce the risk of error, ultimately improving the quality of research and resulting products.

The high-level gap analysis map shown in Figure 2 depicts where existing research infrastructure provides varying levels of support for the roadmap vision. While we found some good existing experimentation tools that can be shared, they only provide a small portion of what is needed to achieve the roadmap's vision. It is important to note that domains are broad and encompass many industries, e.g., transportation includes air, automotive, rail, etc. While one industry may have advanced capabilities, other industries in the same domain may lag behind. Finally, many industries may have modeling and simulation and other advanced testing capabilities (e.g., the University of Michigan's Mcity for testing driverless cars [13]), however the focus of such tools is not cybersecurity. Some of these tools may also be useful as a foundation for building domain or industry specific cybersecurity experimentation tools, but they alone are not sufficient.

The following section discusses the CEF roadmap to produce a shared, vetted, and reusable set of tools and methodologies to catalyze rapid, rigorous cybersecurity research.

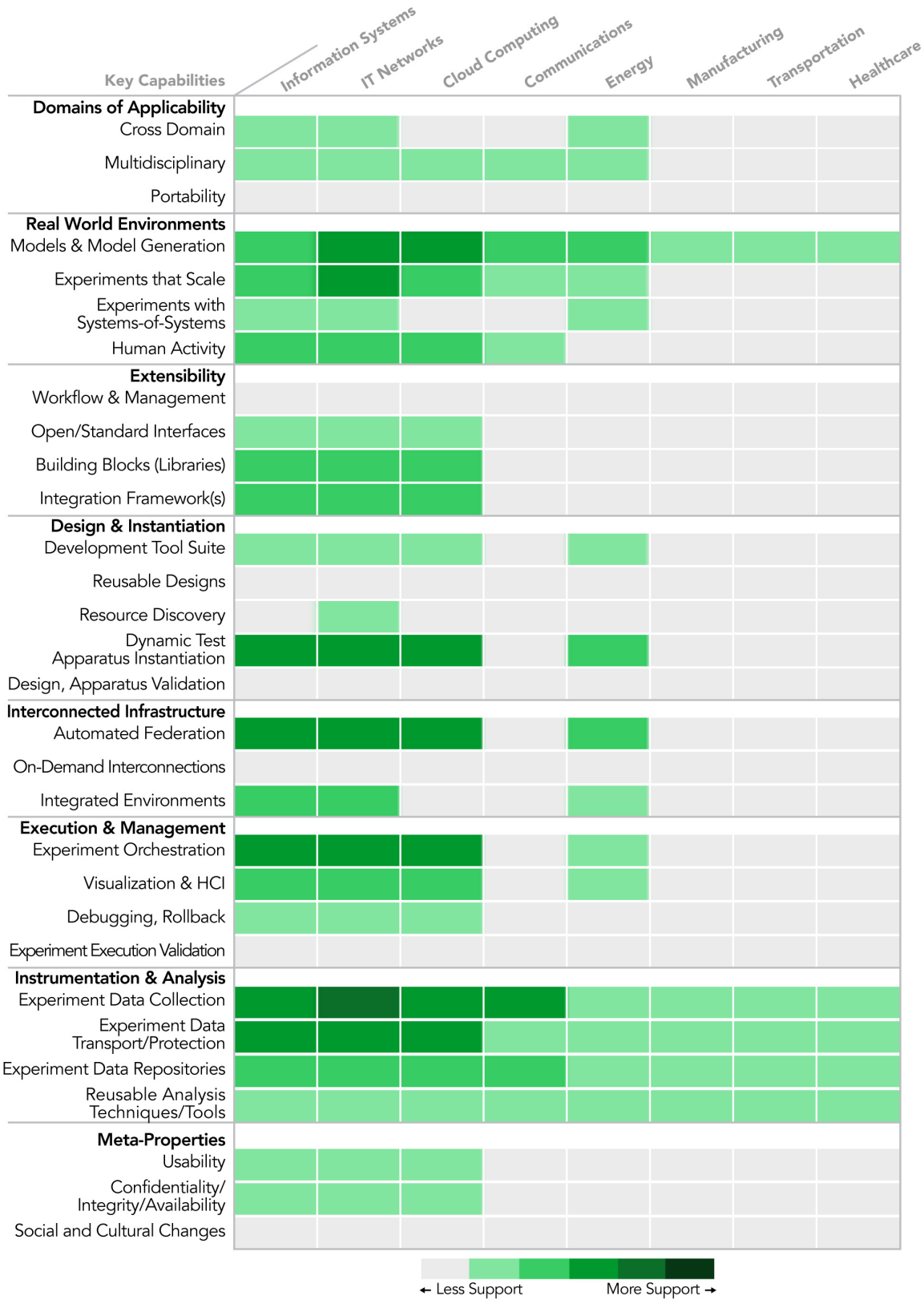


Figure 2. Current research infrastructure support for the needed key cybersecurity experimentation capabilities, with regard to testbeds, tools, and methodologies that are openly available to the broader research community.

4 ROADMAP FOR EXECUTING THE CEF VISION

As described earlier, our key finding is that transformational progress in three distinct, yet synergistic, areas is required to achieve generational advances in the field of experimental cybersecurity research.

Here we present a roadmap for pursing research and development to help research, design, and ultimately build the cybersecurity experimentation infrastructures needed to support these advances. The CEF study identified 30 key capabilities in eight core areas that are required as part of future research infrastructures. The roadmap presents requirements, objectives and goals for these capabilities over identified three, five and ten year phases. In some case the phases build upon each other and in other cases new fundamental research is required over a long period of time to satisfy the objectives of the roadmap.

The capability areas are organized in a layered structure from the outside "application" layer down to the base system and a corresponding set of meta-properties. Thus they move from domains of applicability, to models, to frameworks, to design, to interconnection, to execution, and finally to instrumentation and analysis. Each of the capability areas and their corresponding capabilities are treated in detail in Section 4.1 to Section 4.8, as shown in Table 1.

Each of the sections follows a similar format. To understand the future vision, the current state, and how to achieve the vision, we ask and answer the following questions:

- *Initial description of the capability*
- *Where do we want to be in the future? What will solutions look like? What will the solutions enable us to do?*
- *Where are we today? What are the current technology and research? What are the important technology gaps and research problems?*
- *What will it take to get us there? How can we divide the problem space? What can be done in the near (1-3 years), mid (3-5 years), and long term (5-10 years)? What resources are needed?*

The research time frames refer to expected time periods for achieving identified milestones.

Near-term Research: Near-term research will tend to be more applied in nature and leverage, integrate, and mature existing tools and technologies, leading to new capabilities that can be readily deployed and used in the near term. Near-term research will also uncover challenges and issues and identify new work and research activities to be undertaken in later years. For example in the area of capturing human behavior near term research might focus on simulation capabilities augmented to create realistic human behavior simulation for specialized domains. Near-term research is expected to produce results in the next 1-3 years. Research activities must already be under way or initiated in the next year.

Table 1. The CEF roadmap identifies 30 key capabilities organized into eight capability areas.

Section and Area	Key Capabilities
4.1 Domains of applicability	Support for cross domain experimentation (critical infrastructure sectors)
	Multidisciplinary experimentation that includes computer science, engineering, mathematics, modeling, human behavior, sociology, economics, and education
	Portability of experiments, packaged for sharing and use in cross-discipline experiments
4.2 Modeling the real world for scientifically sound experiments	Models of real world environments
	Experiments that scale
	Experimentation with systems-of-systems
	Human activity
4.3 Frameworks and building blocks for extensibility	Workflow & management (comprehensive, human)
	Open/standard interfaces (API for extensibility, plugins write to API)
	Building Blocks (libraries)
	Tool integration framework (to glue pieces together)
4.4 Experiment design and instantiation	Design tools, specifications, ontologies, compiler
	Reusable designs for science-based hypothesis testing
	Automated discovery of local and distributed resources
	Dynamic instantiation of domain-specific test apparatus
	Validation of instantiated test environments and apparatus
4.5 Interconnected research infrastructure	Automated, transparent federation to interconnect resources
	Dynamic and on demand, with sharing models
	Support integrated experiments that include real, emulated (virtual), and simulations
4.6 Experiment execution and management	Experiment orchestration
	Visualization and interaction with experiment process
	Experiment debugging with checkpoint and rollback
	Experiment execution validation
4.7 Instrumentation and experiment analysis	Instrumentation and data collectors
	Transport and protection mechanisms
	Data repositories
	Data analysis
4.8 Meta-properties	Usability (experiments, owner/operator)
	Confidentiality, availability and integrity of experiment ecosystem
	Social and cultural changes

Mid-term Research: Mid-term research will also tend to be more applied in nature and extend the results of near-term activities. Mid-term activities may increase the scale or complexity of earlier work, or apply work performed in one domain (e.g., the information technology domain) to other domains as part of the goal of multi-domain capabilities. mid-term research may feed off of challenges and issues emerging from near-term activities, and will not rely on solving fundamental research problems that require longer-term activities. For example in the area of experiment design and instantiation mid term research could

include definitions of ranges of repeatable and reusable experiments along with formal specification of degrees of reusability. Mid-term research is expected to produce results in the next 3-5 years. Research activities must generally be initiated in the next 1-3 years.

Long-term Research: Long-term research will tend to address fundamental, theoretical hard problems in computer science, engineering, social sciences, and other disciplines. While advances in these areas are needed to advance CEF, they are not necessarily exclusive to CEF. However, addressing these problems may be easier in the context of CEF rather than the broader context. Examples areas for long-term research include design specification and ontologies, reasoning and semantic analysis, and automated resource discovery. Examples in the area of experiment design include theoretical work in definition of non-interference for experimentation and models of correctness and experiment validity. Long-term research is expected to produce results in the next 5-10 years including new technologies and tools that can feed new, future capabilities. Long-term research activities must generally be initiated in the next 1-5 years.

CEF Ecosystem and Hybrid Architectures

The goal of the roadmap capabilities is not to create a single instance of a cyber experimentation facility or a static definition of a testbed rather taken together these paint a vision for a new generation of experimental cybersecurity research that offers powerful assistance towards shifting the asymmetric cyberspace context to one of greater planning, preparedness and higher assurance fielded solutions.

It is expected that over time the roadmap may be realized through an ecosystem of many different instantiations, from small stand-alone, to localized, to large distributed experimental capabilities.

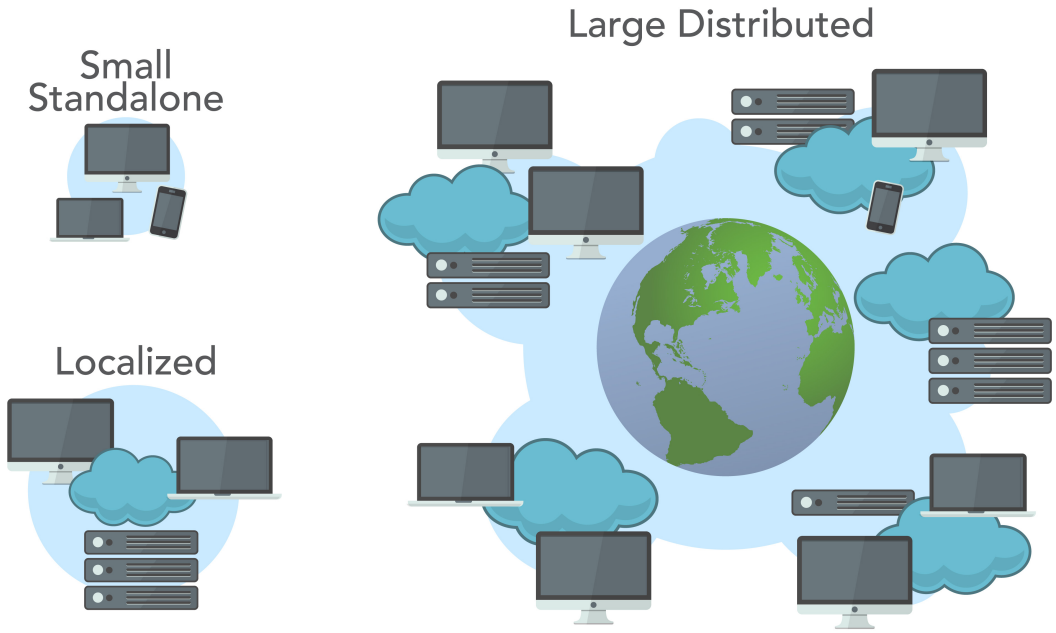


Figure 3. Over time the CEF roadmap may be realized through an ecosystem of many different instantiations – from small stand-alone, to localized, to large distributed experimental capabilities.

The best value will be realized through an ever-changing collection of hybrid architectures that today could consist of combinations of cloud technology, software defined networking, and real server class nodes, as well as emulated and simulated environments and specialized hardware. In conjunction with these components will be approaches to knowledge sharing and community environment via integrated development environments. Thus, just as there is no intention of a single instance of a cybersecurity experimentation facility or a static definition of a testbed, there is no expectation of a single hardware or software substrate.

4.1 DOMAINS OF APPLICABILITY

Initial Description of the Capability

Society's computing dependencies are rapidly evolving. In nearly every aspect of our lives, we are moving toward pervasive embedded computing with a fundamental shift in network properties. These changes bring a very real and wide-ranging set of challenging cyber threats with significant implications across critical sectors of our society. In order to address this situation cyber experimentation needs to be applicable across multiple domains and communities. Cybersecurity is no longer the sole purview of computer scientists but needs to also be accessible to researchers in critical infrastructure domains such as energy, transportation, manufacturing, finance, healthcare, economics, human behavior and many others.

Figure 4. Cybersecurity experimentation spans multiple domains and multiple disciplines.

Furthermore support for packaging and portability of experiments contributes to cross-discipline experiments. The general requirement for reusable designs for experiments is discussed in Section 4.4. Here the requirement is focused on portability particularly targeted towards providing experimental components for non-computer science domains to ease the learning curve for specialized experimentation.

Thus, an important requirement is that future cyber experimentation capabilities support a wide range of communities and a variety of usage models. This manifests in the following three requirements:

1) Support for cross domain experimentation
2) Multidisciplinary experimentation that includes computer science, engineering, mathematics, modeling, human behavior, sociology, economics, and education.
3) Portability of experiments, packaged for sharing and use in cross-discipline experiments

Taken together these capabilities provide research infrastructure supportive of a wide range of users and use cases.

Where do we want to be in the future? What will solutions look like? What will the solutions enable us to do?

In the future a researcher investigating cybersecurity threats and solutions in CIP and other domains, will be able to use specialized research infrastructure coupled with tools and methods to analyze the interplay of domain technologies, human factor models and economic factors. Non computer-scientists will be able to take advantage of the common underlying frameworks as described in Section 4.3. In addition to the framework, there will be support for domain specific light weight specialized instantiations, which will allow for new experimentation to explore interdependencies across domains.

In addition to frameworks and specialized experiments for multiple domains of investigation, new research infrastructure must include the ability to capture, model and recreate human behavior and integrate real humans into experiments (see Section 4.2). The inclusion of models and real human behavior in experiments is important across all aspects of CEF, but will likely have an even more important role in new domain sectors. Initial work is beginning to demonstrate capabilities in modeling cyber-human interactions in nuclear power plants, hospitals, and home computing scenarios.

Finally, integration across multiple disciplines will be enhanced through creation of communities of multidisciplinary researchers, through conferences, publications and cultural changes encouraging a practice of collaboration and sharing (see Section 4.8).

Table 2. Summary of Current State, Vision, and Needed Research for Domains of Applicability.

Capability	Current State	Vision	Needed Research
Cross domain	Limited, mostly stand-alone, sector specific infrastructure for cybersecurity experimentation	Common underlying frameworks, light weight specialized instantiations, emerging pair wise, multi-point ability to explore interdependencies	Common framework, open interfaces, specialized components, modeling of interdependencies
Multidisciplinary	Emerging area of study and research infrastructure; see early research at ARL CRA, Illinois SoS Lablet, Indiana CUTS; early research but little infrastructure support	Ability to capture model, recreate human behavior and integrate real humans into experiments (see 4.2)	Fundamental research across fields to define new multidisciplinary cybersecurity topic, new experimental methods and tools for integration
Portability of experiments	Very little sharing and thus few to no approaches. Requires extensive work to reach across disciplines. Some work at centers, e.g., ARL CRA, Illinois TCIPG, UC Berkeley FORCES	Common practice sharing, community of multi-discipline researchers, conferences, publications, cultural changes	Incentive and social-cultural efforts to change mode of operation, need for creating fields

Where are we today? What are the current technology and research? What are the important technology gaps and research problems?

Currently, most cybersecurity research infrastructure includes limited support for cross domain experimentation. There are a number of mostly stand-alone sector specific testbeds. Most of these have limited support for research experimentation and exploration of inter-dependencies between cyber and the specific sector. In addition these are often closed facilities not available to the larger research community.

There is some evidence of new multidisciplinary research programs in cybersecurity along with the use of research infrastructure. For example funded research from the NSF, Army Research Lab (ARL) Collaborative Research Alliance (CRA) [16][17], University of Illinois Science of Security (SoS) Lablet [18], and University of Indiana Coordinating User and Technical Security (CUTS) project [19], and new work in human behavior and cybersecurity. While these programs are pursuing topics in multidiscipline research in cybersecurity they often have limited support for developing research infrastructure.

While we strongly believe that portability of experiments (see Section 4.3) is an important factor in encouraging multidisciplinary research there is currently very little in practice.

All of these areas require extensive work to reach across disciplines and to create research infrastructure that can support experimentation across multiple domains and encourage sharing. The gaps are both in technology to assist in specializing research infrastructure for cross-domain experimentation, and in culture and common practices around sharing and building multi-discipline communities.

Table 3. Research Milestones for Domains of Applicability.

Capability	Near-term	Mid-term	Long-term
Cross domain	Models and tutorials for non-CS researchers to use RI. Worked examples of multi domain experiments	Collections of specialized components across multiple domains, cultural changes	Common framework, open interfaces, specialized components, modeling of interdependencies
Multidisciplinary (A Top 5 Recommendation)	Larger than RI, needs community cultural changes to bring different disciplines. Incentive and social-cultural efforts to change mode of operation, need for creating fields	Tools and methods for multidisciplinary experimentation, e.g., power engineer models or psychologist inputs to human behavior models Present a user experience tailored to the domain and community of interest	Fundamental research across fields to define new multidisciplinary cybersecurity topic, new experimental methods and tools for integration
Portability of experiments	Depends on common API/interfaces (Section 4.3), build models with well defined interfaces, isolate domain specific components for clean interaction with RI	Composition tools to integrate domain specific models	Automatic encapsulation and compilation of experiments for use across multiple environments

What will it take to get us there? How can we divide the problem space? What can be done in the short (3 years), mid (5 years), and long term (10 years)? What resources are needed?

Near Term – 3 Years

In order for new experimental capabilities to have the broadest impact, it will be necessary to lower the barrier to entry for researchers and users from non-computer science and networking communities. In the near term a number of both technical and social/cultural activities should be undertaken.

There are a number of initial efforts in this area that should be leveraged, some are directly in the area of cybersecurity analysis and experimentation for specialized domains while others are building up multi-discipline research and development in areas of critical infrastructure protection. Notable efforts can be found in the TCIPG Project lead by University of Illinois [20] and the NSF Frontier projects in Cyber Physical Systems, Foundations of Resilient Cyber-Physical Systems (FORCES) led by UC Berkeley [21], and Correct-by-Design Control Software Synthesis for Highly Dynamic Systems led by University of Michigan [22].

To begin with it will be important to create worked examples of cybersecurity experiments from multiple domains. Through efforts at USC-ISI, Pacific Northwest National Laboratory, Illinois, UC Berkeley and other organizations there is a growing body of research and experimentation in cybersecurity of power systems (largely related to Smart Grids). A catalog of such experiments and the supporting models, data, and analyses should be developed. In addition to cataloging of current multi-domain experiments, tutorials should be developed with a particular orientation for non-computer science researchers. It should be noted that cataloging and developing tutorials will, in and of itself begin a socialization process as researchers from other domains will need to curate their experiments for inclusion in the catalog and assist with developing tutorials. It should be noted that this should not be a one time activity, rather work should be devoted to creating cataloging approaches and increasingly sophisticated tutorials over time.

In addition to a straightforward development of catalogs and tutorials, supporting cybersecurity experimentation across multiple domains will be enhanced through the use of common API/interfaces as described in Section 4.3. These will allow researchers to build specialized models that can use underlying cybersecurity experimentation facilities via the well-defined interfaces. Architecturally it will be important to isolate domain specific components from experimentation engines in order to design clean interaction with the research infrastructure. Along with this, common API's will allow domain specific experiment methods, tools and UI's to be developed. The more domain specific (e.g., energy, transportation, financial services) interfaces and terminology that can be used the easier it will be to expand the community of users.

In many ways more than the development of various research infrastructure components, will be the need for community and cultural changes to bring different disciplines together. There will need to be incentives and social-cultural efforts to change current modes of operation, and to create new cross discipline fields. We have seen such changes begin to take hold in various science disciplines with increased focus on data sharing, analysis and

the use of metadata. While cultural changes take time, it is important to initiate them early in the process.

Mid Term – 5 Years

While the near term activities will focus on catalogs and tutorials around experiments conducted by a small set of pioneering multi domain researchers, the mid term should work to create collections of specialized components across multiple domains. These components can facilitate cross discipline experimentation. As our critical infrastructure is increasingly dependent and integrated with communication systems it will be important to quickly and easily encapsulate technologies for experimentation. However it should be noted that this is not simply a packaging problem. Misapplication of experimental components can lead to wrong or misguided experiments that may yield erroneous results.

Thus, it is important to begin work (some of which will be conducted under the 10 year roadmap activities) on methods for establishing correctness of experiment components. These may include approaches to specifying and reasoning about constraints, model description frameworks and structured guidance for application of cross domain experimental components against interdependent research problems. For example a researcher in electrical power systems wants to study the effect of distributed communication as part of a Remedial Action Schemes for responding to transmission overload. Conducting such an experiment requires both models of the power systems in question, and communication systems. Each discipline, power engineering and distributed communications have well understood and defined models of behavior and methods for experimentally varying and validating their models and implementations. However if care is not taken in composing the multiple models then invalid experimental results could occur.

Verification and validation of models and the interplay of models and experimental systems under test is an area of research and development that fits within the mid term research roadmap. This is also an example of an area that can draw from, and benefit from, long standing and on-going research in the computer science community as well as in each of the sub disciplines.

In parallel with research and development of a host of challenges in experiment validation and validation/verification of experimental components, is the need to develop methodological resources for the experimenter. Further these methodological resources should be realized in tool chains for researchers. Developing tools chains that bridge different domains will have the greatest impact on enabling multi domain and multi discipline cybersecurity experimentation.

Finally another area to help a wide range of researchers to rapidly work in an experimental facility is to develop methods to present user experiences tailored to the skills and background of the user communities. Crafting tailored user experiences is not a small task and requires adoption of the terminology, workflow and methods traditionally preferred within an established community-of-interest.

Long Term – 10 Years

In the long term, progress needs to be made on the wide range of fundamental computer science and information technology topics that underlie the vision of common frameworks, open interfaces, specialized components, and modeling of interdependencies.

Layered on top of these fundamental research topics there should be activities focused on tailoring and applying the general research to the field of science based cybersecurity experimentation and test.

In addition to fundamental advances in underlying concepts and technology, there should also be efforts focused on fundamental research across fields to define new multidisciplinary cybersecurity topics, and new experimental methods and tools for integration of these disciplines.

Finally the further out the research agenda looks one should expect to find ever increasing automatic encapsulation and compilation of experiments for use across multiple environments.

4.2 MODELING THE REAL WORLD FOR SCIENTIFICALLY SOUND EXPERIMENTS

Initial Description of the Capability

For research in cybersecurity to be impactful, it must be based on both sound science and on the real world. The community needs shared, validated models and tools that help researchers rapidly design meaningful experiments and test environments. These models are needed for both real and simulated test environments. The problem space may be divided into four subcategories:

1) Models of real world environments
2) Experiments that scale
3) Experimentation with systems-of-systems
4) Human activity

This set of capabilities is what grounds research, making sure it useful in solving real world problems.

Figure 5. Cybersecurity experimentation must capture real world scale and human activity.

Where do we want to be in the future? What will solutions look like? What will the solutions enable us to do?

Researchers across all domains that rely on computational systems will rapidly design and run experiments that reflect the real world by reusing and extending validated, real world models of systems, actors, and behaviors. This will increase the probability that experimental results will solve real world cybersecurity problems.

A robust set of tools and ontologies are needed to automatically or semi-automatically extract models of domain-specific computational devices, of both simple and complex networked environments, and of system actors, such as benign users, attackers, and administrators. A set of rich ontologies is needed to be able to describe the diverse features and behaviors that exist across many domains. It is infeasible and unnecessary to model all details for all experiments. However, a lack of fidelity where fidelity really matters can make an experiment meaningless. Techniques are needed to help researchers identify the levels of fidelity that are needed and where. When extracting and extending models, researchers will specify the fidelity and salient features to be captured.

The use of flawed or invalid models can invalidate experimental results, and therefore, models must be validated. Techniques and tools to automatically analyze models for validity are needed. Once validated, researchers will use these models as they design and instantiate experiment environments (see Section 4.4).

Given the trend towards larger and more complex interconnected systems, researchers will necessarily design and conduct experiments at scale. The need for varying levels of fidelity in system representation is especially important when conducting research on large, complex systems-of-systems. For example, modeling the entire future smart transportation system in great detail is not feasible, not only because of the amount of details that must be captured, but also because it will not be a static "thing". Rather, it will be a rapidly changing

environment. Modeling the system's salient features, including its dynamic nature, is feasible.

Usually when one thinks of scale, he thinks in terms of the size of the network. When considering levels of model fidelity, however, one sees that scale is multi-dimensional. Depending on the area being explored, a researcher might want to consider scale in the number and/or types of nodes, number of routers and switches, the number of operating systems represented, the number of users and/or user types represented, the amount of time, the amount of storage and processing, etc. Researchers will specify the scale for all dimensions as part of their experiment design.

Researcher will use smaller device abstractions and system models as building blocks to compose larger systems-of-systems models, without having to specify every single device. Consider a low-resolution photograph: when rendered in the large, blurry areas appear where there is no data. A tool that is able to extrapolate and smartly generate realistic data to fill in the gaps is desirable. The same is true for designing large system-of-systems models from smaller components. When the design is realized in a test environment, system model will automatically scale without losing resolution or fidelity.

Researchers will, within the test environment, accurately represent complex human behaviors, including those of diverse individuals, groups of individuals, and large populations. Realistic human behavior extends well beyond simply browsing web sites, fetching and sending email, and editing Microsoft Word documents. Humans frequently work in groups to accomplish goals. They engage in byzantine and reactive behaviors, modifying their behavior based on changes in the cyber environment, external real world events, time of day, stress, fatigue, and other factors. Further, human behavior is based on different experiences and levels of knowledge. Behavioral research is needed to understand how economic incentives and other factors impact decisions and how humans learn and adapt over time. Research is also needed to form a theoretical understanding of what must be included in human behavior models for them to be "realistic enough".

Humans also interact with compute devices outside of the IT domain (e.g., automobiles, embedded medical devices, etc.). Human-computer interaction in other domains differs significantly from the use of personal computers, laptops, and mobile devices. These domain-specific behaviors must also be modeled and available for use by researchers. Behavioral research is needed to understand cross-domain human behaviors. Finally, similar to system models, these models of behavior must also be validated prior to use in experiment designs (see Section 4.4), so such techniques need to be developed.

When using real humans in an experiment, a means to describe and enforce the rules of engagement is needed. Techniques are needed to monitor human activity and automatically extract an accurate model for use in subsequent experiments. Finally, the introduction of humans in experiments, either as an experiment subject or to create behavioral models for use in experiments, raises human subject research issues that must be addressed. The Menlo Report [23][24], funded by the Department of Homeland Security Science and Technology Directorate, discusses the ethical issues of human subject research in cybersecurity experimentation.

Table 4. Summary of Current State, Vision, and Needed Research for Modeling the Real World.

Capability	Current State	Vision	Needed Research
Models of real world environments	Manually constructed environments, based on human understanding of real world; fidelity is limited	(Semi-)automatically extract models of real computational devices and environments; automated model validation against real device or environment; rapid injection of models into simulation and physical test environments	Semantically rich language to model complex computing environments across diverse domains; techniques to extract real world models for test environments, model validation
Experiments that scale	Scale via simulation at fixed fidelity is norm	Multiple dimensions of scale, ability to stretch experiment without affecting fidelity	Scaling, models of complex systems-of-systems, fidelity and building blocks
Experimentation with systems-of-systems	Little to no advanced experimentation; perhaps opportunity to leverage T&E community	Ability to experiment via abstractions and to reason about composition	Abstractions, representation, multiple dimensions of scale, validity
Human activity	Several stand alone tools that are on host and based on statistical models of human behavior; at least one non-interfering off host tool exists that drives the user interface, playing the role of a human; at least one solution for virtualization non-interference	Ability to accurately represent fully reactionary complex human and group activity in experiments; include live and synthetic humans without artificialities that may interfere in some experiments; capabilities to help ensure scientific validity when including live humans in experiments	Theoretical understanding of model contents needed for realism; automated creation of human models based on observations; wide range of behavioral research to understand decision process in different domains; models for non-IT domains; human activity at scale

Where are we today? What are the current technology and research? What are the important technology gaps and research problems?

Experiment environments today are manually constructed based on the researcher's understanding of the real world, and often have limited or incorrect fidelity. The models represented in these environments are not rigorously validated for real world applicability, and this is an area of significant deficiency in cyber experimentation. Given that solutions are to work in the real world, creating and using valid, realistic models in safe, controlled, laboratory environments are critical to cybersecurity research. At present, however, there are fundamental research barriers to this goal. A semantically rich modeling language is required for modeling complex and sometimes inter-domain computing environments.

Techniques are needed to observe a real world environment, automatically extract the salient features, create a valid, real world model, and then inject into a test environments. Techniques are also needed to validate handcrafted models.

Very limited capability exists to perform large-scale experiments. Such can be difficult and costly to implement in test environments and thus are often performed via simulations of a fixed fidelity. Some simulation capability exists for large-scale networking (e.g., OPNET[25]). Amazon's Mechanical Turk [26] is an example of a non-simulation capability that has been used for large-scale cybersecurity based user studies. Additional research is needed, especially to determine ways to achieve fidelity at scale.

Little support exists for composing and experimenting with systems-of-systems. The systems engineering community has test capability for systems-of-systems and may provide a starting point for an early capability. Research is needed to understand how to create valid system abstractions and multi-dimensional representations to support experiments at scale.

A moderate amount of on-host human activity emulators exist for use in the IT domain. The MIT Lincoln Laboratory (MIT-LL) LARIAT tool [27]is based on statistical models of human behavior. It requires an agent on each host that interacts with the operating system and applications on behalf of a user. For some experiments, this is fine, however such could interfere with other experiments. MIT-LL recently developed a companion tool that is less intrusive called K0ALA [28]. It runs off-host and emulates a human by using image recognition to "read" VGA output and then injects keystrokes and mouse clicks via USB port. The MIT-LL technologies are limited in use for the U.S. government and government contractors only. Skaion Corporation [29] has a "for fee" human activity generation tool that is available to anyone who has research budget to cover the costs. USC-ISI more recently has invested in high fidelity models of human behavior through their DETER Agents Simulating Humans (DASH) project [30], where they examined the effects of diversions and fatigue, among other things, on human behavior.

Existing and emerging synthetic human activity technologies and models do not fully meet the CEF vision but should provide a basis to reach a large part of that vision in the near-term. Simulators in other domains such as air and traffic may also be useful as foundations for CEF purposes.

Moving forward, research must be conducted to understand what must be included in models to provide the level of realistic human behavior needed for real world relevant experimentation. Techniques are needed to instrument and observe the behavior of real humans and to automatically extract valid behavioral models, and then inject into a test environments. In addition, behavioral research is needed to understand the human decision process in different domains of interest, such as transportation, medicine, and energy. Finally, research is needed to understand human behaviors in the context of multiple communities, including collaborative efforts and more general collective group behaviors at scale. This research will necessarily involve cybersecurity and domain experts and human cognition scientists. Techniques are also needed to validate these models.

Table 5. Research Milestones for Modeling the Real World.

Capability	Near-term	Mid-term	Long-term
Models of real world environments	Transition semi-automated model creation capability from government, university, and industry labs	Automated model creation for IT domain; semi-automated model creation for other domains (e.g., transportation, electric power)	Automated model creation for other domains (e.g., transportation, electric power);
Experiments that scale	Multiple scaling mechanisms, virtualization and models	Scalable infrastructure for experiment control and management, real-time agents for rapid control of large scale experiments	Extreme scale, hybrid emulation and simulation experiment artifacts (validated)
Experimentation with systems-of-systems	Frameworks for abstracting system level structures. Use of test harness technology from system engineering communities	Experiment definitions that build on integrated components along with experiment management and data analysis at co-joined systems	Methods for reasoning about complex interactions between composed systems
Synthetic human activity (A Top 5 Recommendation)	Transition existing human activity simulation capabilities and augment to create realistic human behavior simulation for IT domain – support human attacker, defender, and user simulation	Accurate, non-interfering human activity simulation for other domains (e.g., transportation, electric power); ability to specify behavioral bounds for repeatability; tools to help ensure scientific validity when including live humans in experiments; assisted extraction of real human behavior, capture into models for subsequent simulation	Automated extraction of real human behavior, capture into models for subsequent simulation. Extrapolation to generate projected human behavior models reflecting changes in technology and human interaction patterns

What will it take to get us there? How can we divide the problem space? What can be done in the short (3 years), mid (5 years), and long term (10 years)? What resources are needed?

Near Term – 3 Years

In the near term, semi-automated IT network model creation capability from national labs and other sources should be transitioned to general use in order to provide an interim capability.

Multiple scaling mechanisms, virtualization, and models should be developed as a first order capability for experiments that scale. This must include the ability to create and experiment with systems-of-systems. Frameworks are needed for abstracting system level structures. It may be feasible to adopt the test harness technology used by the system engineering communities. This should be explored.

A fairly robust, non-interfering synthetic human activity capability for the IT domain may be achieved in the near term by leveraging and integrating the USC-ISI DASH advanced human behavior models [30]with the LARIAT [27] and Skaion [29] traffic generation frameworks. MIT-LL's K0ALA may be used to provide off-host non-interfering activity simulation for physical and virtual hosts.

Most existing behavioral models focus on IT users. Realistic attacker and defender models should be developed as well. It may be feasible to borrow partial models from computer-based strategy games. Another potential source of knowledge is red teams that study different types of adversaries for the purpose of replicating them in cyber defense exercises. One such red team is the Sandia National Laboratories Information Design Assurance Red Teams (IDART) [31]. The resulting synthetic human models should include diverse, byzantine models of goal-oriented human behavior, to include actions and reactions and group/community-based interactions.

Mid Term – 5 Years

In the mid-term, effort should be expended to provide a first order model creation capability for non-traditional IT domains and also in expanding and automating the model creation capability for the IT domain. The semi-automated model creation capability for IT networks should be transitioned to all other domains (e.g., energy, medical, transportation) providing a first order capability. This will necessarily require an extension of the model ontology to include concepts from these other domains.

For the IT domain, efforts should focus on providing the ability to automatically extract the salient features of an IT network and create models of those networks that may then be injected into a test environment and instantiated as part of a real world scenario. This capability serves to provide the realistic foundation needed so that research results are relevant and useful when transitioned into operational use.

Scalable infrastructure for experiment control and management should be developed, inclusive of agents that are capable of, in real-time, modifying the execution of a large scale experiment based on either direct researcher interaction or pre-planned trigger events.

In support of experimentation with systems-of-systems, the capability to define experiments that build on integrated components from multiple-domains should be developed. In addition, experiment management and data analysis must be expanded to support experiments with co-joined systems.

Accurate, non-interfering human activity simulation should be developed for other domains (e.g., transportation, energy). This will require research to understand how humans use these systems and the wide range of behavioral models necessary to simulate humans in these domains.

For the IT domain, synthetic human activity generators should be enhanced to support the specification of behavioral bounds for the activity generated in order to enable repeatability

of experiments. The off-host synthetic human activity generation capability should be expanded for scalable experiments on physical hosts.

Finally, tools should be developed to support the inclusion of live humans in scientifically valid experiments. These instrumentation tools are needed to capture and measure the performance of humans and their interaction with the system. Tools should be provided to assist researchers in converting captured human activity into human behavioral models for subsequent simulations.

Long Term – 10 Years

In the long term, efforts should focus on providing the ability to automatically extract the salient features of other domains (e.g., transportation, medical, and energy) and create models of those networks that may then be injected into a test environment and instantiated as part of a real world, domain-specific scenario. This capability serves to provide the realistic foundation needed so that research results are relevant and useful when transitioned into operational use.

In the long term, the experimentation capability must support experiments at the extreme scale, using a hybrid of validated emulation and simulation artifacts. In addition, the composition of systems will create complex interactions between systems, so methods are necessary to reason about these interactions.

Automated tools should be provided to convert captured human activity into human behavioral models for subsequent simulations. Finally, tools should be able to generate human behavioral models based on projected human interaction patterns with technology as it changes in the future.

4.3 FRAMEWORKS AND BUILDING BLOCKS FOR EXTENSIBILITY

Initial Description of the Capability

Frameworks for cyber-research facilities are required to meet the CEF vision for cybersecurity experimentation using infrastructure that is open, accessible, covers the entire lifecycle of experimentation, and is extensible and applicable to multiple domains of research as described in Section 4.1 on the Domains of Applicability capability.

Frameworks in this context are composed of architectures, components, design patterns for integration, and for instantiation of a cyber experimentation facilities. Components will be developed for specific goals of a framework, including workflow and administration, and composition and specialization. Architectures will provide overarching structure for integration and instantiation. Integration is the use of components to put together the technology for operating a particular facility for cyber-experimentation. Instantiation is the creation of a specific facility by both integrating components and specializing for the research targets of a particular community or domain of research that the facility will serve.

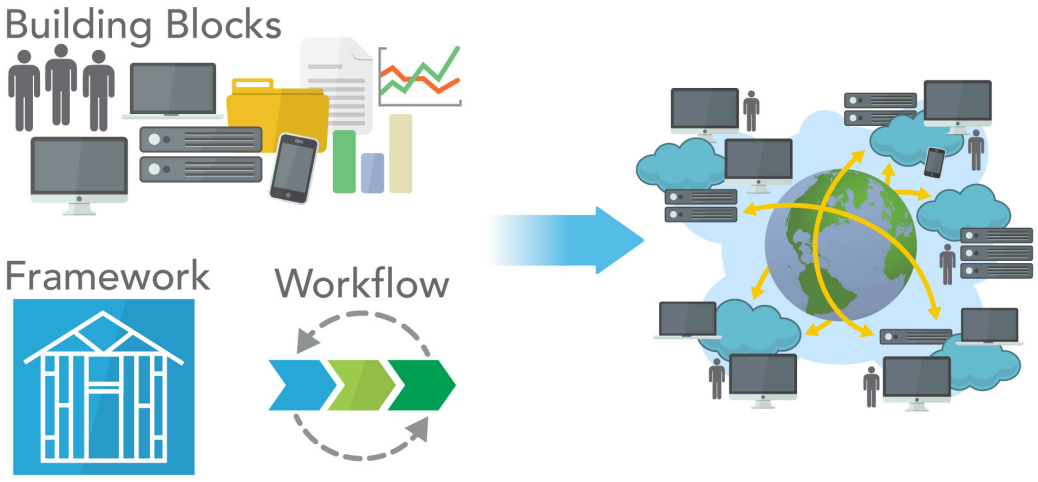


Figure 6. Workflows, building blocks, and frameworks to build extensible experimentation infrastructure.

In a framework for extensibility, there are three related aspects, as well as a fourth area for integrating them:

1) Workflow and management of experiment lifecycle that includes both comprehensive management capabilities and adaptable human interfaces for using these capabilities.
2) Open interface APIs for these research infrastructure capabilities that support both extensibility/specialization of human interfaces, and plug-in architectures for extending the capabilities of a particular instance of a facility.
3) Libraries and other building blocks both for specialization/extension of facilities, and composition both at the level of experiments and at the level of facilities.
4) Integration frameworks, tools, and designs to compose the above three aspects into infrastructure components and particular instantiations into domain-specific facilities.

Where do we want to be in the future? What will solutions look like? What will the solutions enable us to do?

The CEF vision includes the development and use of wide scale, multi-discipline, multi-institution experiment research infrastructure that can be loosely or tightly coupled via largely dynamic and automated methods. Achieving this vision requires substantial research and development into the enabling technology for cyber-research facilities, so that future facilities can operate with the above characteristics, and be extensible, specializable, and highly usable. In each of the three areas identified above, and the fourth area of integration, the sought endpoint of this research can be described as follows.

Workflow refers to the process for using research-infrastructure technology whereby an experimenter (CEF user) conducts experimentation. Workflow related technology includes:

- The user interface for guiding an experimenter through a facility's capabilities (actually, multiple user interfaces for multiple audiences, domains, etc.),
- The underlying technology that implements the workflow concepts and processes;
- User interface and underlying technology for non-experiment-related activity such as user profile management, communication, data sharing;
- User interface and underlying technology for administrative workflow for facility operators.

Management includes experiment management and facility management. Experiment management is the set of activities that an experimenter does within a cyber-research facility, to conduct experimentation, including the use of one or more workflow. Facility management is the set of activities that are performed by staff operating a CEF for the use by experimenters.

Although much research should focus on experimenter workflow and experiment management, the scope of research, and the vision for the future resulting form it, includes all of the infrastructure technology for a cyber-research facility.

Open and standardized interfaces will be a key feature of the enabling technology for these facilities. The common model of experiment lifecycle will be implemented using components and APIs. As a result, an instantiation of this technology, as a specific facility, will be extensible in several ways:

- Workflows and user interfaces can be customized and specialized, without perturbing the underlying enabling technology and the APIs of its components.
- Experiment definition is extensible, with a plug-in architecture for additional and/or specialized experiment aspects.
- Alternative or customized tools can be created for specific methods of performing experiment management activity, in discrete phases of the experiment lifecycle.
- Infrastructure and resources will also be extensible via plug-ins to accommodate new types of resources, and cross-facility sharing of them.
- Infrastructure and resource type extensions will benefit from integration via standardized APIs to plug-ins that utilize an extensible common object model for experiment resources and components.

As a result, the enabling technology for a facility should be sufficiently flexible to support domains that are currently beyond the horizon, and the tools, techniques, models, and

processes of experimentation that will have evolved to support these new domains of cyber-research.

Just as there must be standards-based building blocks for building a facility, each facility must have significant infrastructure for experimenters to use and share **building blocks for experiments**. These will include experiment components, infrastructure within experiments, and ontology-driven metadata definitions. In addition, facilities will have meta-building blocks using knowledge bases, provenance tagging, cataloging, searching both for experiment building blocks and for meta-information about methods used and lessons learned by other experimenters.

Taken together, all of these framework elements, components, repositories, etc., will serve to drive **integration** at two levels: experiments and facilities. At the level of cyber-research facilities, instantiation will be performed using worked examples of integration in other facilities, but specialized to the needs, workflows, resources, models, etc., of that facility's user base and domain. At the level of experiments, these facilities will have vastly increased ease of use for experimenters to share, re-use, and adapt not only experiment apparatus and operation, but also domain specific experimenter developed tools contributed to a facility.

Table 6. Summary of Current State, Vision, and Needed Research for Frameworks and Building Blocks for Extensibility.

Capability	Current State	Vision	Needed Research
Workflow & management	Rudimentary support, mostly highly manual	Integrated set of methods, tools and procedures to manage all stages of experimentation life cycle and reduce cognitive	Definition of experiment life cycle, specific methods and tools for different phases in lifecycle, human to infrastructure interfaces, approaches to capturing experimenter intent
Open/standard interfaces	Some standards per facility, e.g., GENI R-Specs, evolving DETER SPI, and MIT-LL, but none across capabilities	Collection of open/standard architectures and interfaces to advance composition and specialization	Defining architectural abstractions that can serve for the long-term Support for extensible domains that are beyond the horizon, tools, techniques, models, and processes that evolve

Capability	Current State	Vision	Needed Research
Building blocks (libraries)	Fairly limited and uniform set of "nodes", "traffic" "connections", and "data"; local and not automated searchable or distributed across the community	Large extensible distributed collections/libraries identifying experiment infrastructure and experiment components, catalogs of lessons learned and domain specific information and rich knowledge base with provenance tagging available for automate search	Cataloging approaches, metadata definitions, ontologies, and knowledge bases Catalog indexing and searching along with catalog consistency mechanisms
Tool integration framework	Limited plug-in capabilities, e.g., GENI, NCR, MIT-LL	Open framework with domain and problem specific plug-ins from which researchers may mix and match to instantiate the needed test apparatus as dictated by the specific research problem being addressed	Common frameworks, tools abstractions, libraries and metadata for tools, usage models and examples, increase ease for experimenter developed tools to be contributed

Where are we today? What are the current technology and research? What are the important technology gaps and research problems?

Workflow and Management: The current state of frameworks for experimentation consists of limited experiment lifecycle models and capabilities, each endemic to a specific testbed or cyber-research facility. Efforts to date, in the area of building experimental infrastructure, have largely been the development of custom systems, e.g., PlanetLab [32], Emulab [33], GENI [3], and NCR [4][5]. Common models and capabilities are rare, and largely limited to inter-testbed federation and related mechanisms for experimentation that spans multiple facilities.

Research into frameworks for experimentation – especially those that are not endemic to a particular facility – could at best be characterized as being in early stages. In the medium to long term, robust experimentation frameworks must support experimenter workflow, including functions for experiment management throughout an entire lifecycle of an experiment. Today, however, experimenter workflow and experiment management are rudimentary and largely manual in practice. Automation of experiment management is part of current research, but at present is more oriented to automation per se, than support for a particular methodology for cyber-experimentation.

Beyond automation per se, the research goal is to create an integrated framework – an extensible set of methods, tools, and procedures – for the management all stages of the experimentation lifecycle. The goal includes not only the comprehensive management capabilities, but also adaptable human interfaces for using these capabilities, where the adaptability enables changes to user experience and workflow, without involving any changes to the underlying capabilities of a cyber-experimentation facility.

Open Interfaces and Standards: Other current research activity includes nascent efforts on standard and/or open definitions that can provide data interoperation or programmatic access to some aspects of the enabling technology for cyber-experimentation facilities, including: system programming interfaces, application programming interfaces, specification languages, and tools for using specification languages. For the results of this research, the vision is two-fold: first, a set of reusable, extensible technologies that enable the repeatable construction of cyber-experimentation facilities; second, the use of standards and open interfaces to customize a newly constructed facility, and extend its capabilities, to meet the needs of a particular community and domains of research. Customization critically includes the specialization of a facility's core experiment support capabilities, to support domain-specific aspects of experiments, lifecycle, workflow, and experiment composition.

At present these interfaces are specific to individual facilities, but are first steps toward defining more general and cross-facility abstractions with multiple benefits sought: common models for experiment and lifecycle; common approaches to composition of experiments, include composition spanning multiple facilities; ability to adapt and specialize the enabling technology of a facility, to meet the needs of a particular community or domain of research.

The overarching challenge of this research is the definition of architectural abstractions that can serve for the long term, while being adaptable and extensible for research domains that are beyond the horizon, with potentially new requirements for experimentation techniques, tools, models, and processes. However, current work suggests some starting points for development of interfaces in a basic architecture that may be relevant to several current or future cyber-experimentation facilities.

Building Blocks: One wide category of building blocks consists of components and interfaces relevant to constructing a cyber-experimentation facility. These building blocks are part of work on interfaces and standards described above. While work on them will be an important part of CEF research, they are almost by definition not building blocks used directly by experimenters, but rather create the abstractions of the building blocks used by experimenters.

By contrast, experimenter-facing building blocks are presented to users of a facility, as capabilities that are used to create and operate experiments themselves.

Current practice is generally low-level and endemic to individual facilities:

1) There is a fairly limited but uniform set of elements such as "nodes," "connections," "traffic," and "data".
2) Each facility has its own specific capabilities to define an experiment in these terms, and the set of terms is not extensible.
3) Re-usability and sharing are possible within a facility, to the extent that an experiment is defined in terms that are portable and copyable, such as a network specification in a specification language.

4) However, such re-usable building blocks may be just as likely to be stored privately by an experimenter, as to be shared with other experimenters. There are limited or no mechanisms within a facility to foster re-use (e.g. searchable repositories), and no mechanism to share between facilities.

In contrast to this current state, the vision for the future is one of: a varied, extensible, growing set of types of building blocks; some standard types in common across many facilities; and rich federated repositories of building block instances.

1) Today's low-level types of building blocks, and the means to define them (e.g. specification languages, data structures, graphical renderings) are standardized for use across many facilities. Higher-level building blocks are used to define larger scale and/or more abstract definitions of experiments' structure, operational procedures, methods of interaction, instrumentation and data collection, and other aspects of experiments that result from ongoing research into methods of cyber-experimentation.
2) The set of types of building blocks is extensible, so that a facility operator can add support for new aspects of experiments, and experimenters use these aspects to define and/or operate experiments in new ways that meet the evolving needs of an experimenter community or domain of cyber-experimentation and research.
3) Experiments, both as a whole, and as individual elements, are stored in libraries, cataloged with searchable structured metadata, unstructured information, lessons learned, and domain specific information.
4) Across multiple facilities (or other repositories), these libraries comprise a large, searchable, extensible, distributed collection of experiment components and experiment infrastructure.

To reach this vision, the research agenda includes:

- Definition of new ways of defining experiments, experiment components, and experiment infrastructure, in a much wider variety of methods that today, for large or varied scale, and large, small, or varied level of detail.
- Extensibility of experiment definition/components/infrastructure to support specific domains and/or communities.
- Standardization of some components (and methods and tools for defining them) that can be inter-operable across facilities.
- Creation within CEFs of approaches to cataloging, tagging, indexing experiments, components, etc., and development of experimenter knowledge bases around libraries constructed using them.
- Development of experiment ontologies and metadata schemae that can define commonalities between re-use models of various facilities, and enabling cross-facility sharing.
- Approaches to structuring cross facility sharing of libraries, including federation of repositories and development of tagging, indexing, and searching methods that span the federation.

Integration: Integration of experimentation workflow and building blocks is an area of nascent research on methods of cyber-experimentation, and research on technologies for constructing and extending facilities for cyber-experimentation. Common interfaces and standards are limited to techniques for operating experiments that are federated across facilities that share a federation model, usually at a low level of abstraction.

In contrast to this very low level of integration today, the future vision is one in which advances in all three areas have created a rich set of workflow capabilities and experiment building blocks, implemented in facilities with open interfaces for extensibility. As a result, both facility operators and experiments will have an open framework with general, problem-specific, and domain-specific plug-ins, from which experimenters may instantiate their needed experimental apparatus and infrastructure, as dictated by the needs of the specific research problem that the experimenters are addressing.

To achieve this vision, much of the research is to be done in the three areas listed above, but the work must be done in a concerted manner; ideally in a holistic manner, but at least with frequent touch points for integration of new capabilities within a given facility, and interoperation between facilities. The above-described common frameworks, usage models, tools, abstractions, examples, libraries should facilitate not only integration, but also extensibility, and increasing the ease with which experimenter-developed advances can be incorporated.

In order to address these challenges new research needs to be conducted into definition of classes of meta-experiment abstractions, design patterns and supporting service and object definitions. In order to achieve wide spread compatibility and adoption it may even be appropriate to engage in some standards process to coalesce specifications and implementations.

Table 7. Research Milestones for Frameworks and Building Blocks for Extensibility.

Capability	Near-term	Mid-term	Long-term
Workflow & management	Definition of experimentation life cycle, survey/adaption of software engineering approaches, tools for different life cycle phases	Increased distributed operation and sharing of tools across distributed infrastructure	Advanced interfaces for experimenter interaction, approaches to capturing experimenter intent
Open/standard interfaces (A Top 5 Recommendation)	Establishment of architectural abstractions, common interfaces, canonical representations	Collection of open/standard architectures and interfaces to advance composition and specialization	Support for extensible domains that are beyond the horizon, tools, techniques, models, and processes that evolve
Building blocks (libraries)	Initial catalog/library of building blocks, extensible with community interface for contributions	Cataloging approaches, metadata definitions, ontologies, knowledge bases	Cataloging approaches, metadata definitions, ontologies, and knowledge bases. Catalog indexing and searching along with catalog consistency mechanisms
Tool integration framework	Common frameworks, tools abstractions, libraries and metadata for tools	Usage models and examples, increase ease for experimenter developed tools to be contributed	Automatic integration, plug and play, interface abstractions

What will it take to get us there? How can we divide the problem space? What can be done in the short (3 years), mid (5 years), and long term (10 years)? What resources are needed?

Near term – 3 Years

Open Interfaces: Research will develop the three foundations of open interfaces:

1) Foundational abstractions, object models, and data models
2) Object and data schemae needed for interoperability of separate components via common open programming interfaces.
3) Common open programming interfaces based on (1) and using (2).

It is unlikely that a single standard set of models and interfaces will emerge, but multiple complementary research efforts may clarify the pros and cons of multiple approaches, and/or segment the total problem into elaboration of separate subsets of a future more comprehensive set of interfaces.

Efforts may be more limited in scope, focused on the more basic abstractions of current experimentation, in order to define re-usable building blocks that implement them.

Workflow: Research will initially focus on one or more rigorous definitions of the life cycle of experimentation, based on exploration of existing approaches and tools for related developmental lifecycles (including but not limited to software development). Multiple research implementations of these lifecycle models will necessarily include the development of new human interfaces of facilities, including user interfaces.

Research into workflow and lifecycle must be enabled by research results in building blocks (see below) especially a new architectural feature in which the human interface to cyber-research facilities can be implemented not as an integral part of a facility (as is typical at present) but rather a separable component that can be modified and explored without any need to change the core capabilities of facility.

With the resulting much more modular and easily extensible human interfaces, research can use the extensibility to explore multiple different approaches to experiment lifecycle implementation. Such exploration must be able to include separable work on specific tools and techniques for separate lifecycle phases, independently assessing various different adaptations of more established methodologies such as software engineering methodologies.

Building Blocks: Three-year phase work in building blocks will include both building blocks of facilities, and building blocks of experiments within them.

For building blocks of experiments, the critical work for the 3-year time frame will be an initial catalog of the existing types of experiment building blocks (e.g. structure, operation, result data capture), and the creation of sharable repositories of libraries of specific experiments and their building blocks. These will include both data interfaces and human interfaces for experimenter community members to contribute to such repositories, and navigate them.

For building blocks of facilities, the critical work for the 3-year time frame is research into architectures for building facilities and extending them. One key focus must be development

of open interfaces that enable separable development of human interfaces as described in the previous section. Another key focus must be separation of the data management components of a facility (e.g. storing experiment definitions, accessing experimental results data) from the core elements for instantiating an experiment and operating it.

With core elements of a facility's building blocks separated from other areas (data management, sharing, user management, user interface), research into extensibility becomes more feasible. Early facilities were limited to one or a few hard-coded building blocks of an experiment, e.g. an experimenter-defined network topology. In the 3-year time frame, the building blocks of a facility's core must become modular to support extensibility as new aspects of experiment definition and operation are discovered in other areas of research.

Integration: Integration will be a critical activity in the latter part of the 3-year time frame, necessary to demonstrate that all the building blocks and open interfaces can put together by instantiating a facility that can support multiple approaches to user experience, workflow, and experimentation methodology. These demonstrations include:

- Capability to instantiate or assemble a facility from a set facility building blocks that implement a set of open interfaces.
- Capability for instantiation to be repeatable and customizable.
- Capability to support multiple approaches to workflow, enabling comparative analysis of them.
- Capability for the workflows to enable repeatable exercises in experimentation that demonstrates the use of experiment building blocks and extensibility.

Mid Term – 5 Years

Workflow: Research in the 4-5 year time frame will leverage the foundation created in the 3-year time frame by cataloging and formalizing experimentation workflows, the tools used in the workflows, and the development of user interfaces to combine them. The 4-5 year time frame will include both extension of that previous work and new focus areas.

The extension of previous work will be the incorporation into workflows and user interfaces the new experiment aspects and new experimentation capabilities developed in the 4-5 time frame and made available by open interfaces (see below).

New areas of research will focus on distributed operation, that is, the operation of experiments that span multiple facilities, and sharing (across a distributed set of experiment infrastructure) of experimentation tools developed in individual facilities.

Open Interfaces: Research in the 4-5 year time frame will be enabled by 3-year time frame advances in architectures for creating extensible facilities, with facility core capabilities carefully separated from UI, workflow, data management, etc. In the 4-5 year time frame, research can focus on the core capabilities of facilities, developing open interfaces and standard architectures for extensibility, e.g., the extension of experiment instantiation to incorporate new aspects – or extend existing ones -- for defining an experiment. Extension is important for specialization of a facility's capability for a particular domain of research; in addition, some domains may require entirely new aspects that must be incorporated into a facility's core capabilities using standard interfaces.

In addition to the focus on extension and specialization of a facility's core capabilities for experiment definition and operation, 4-5 year research will have a similar focus of open-

interface development needed to extend and specialize individual experiments. Experiment composition (creation of a new experiment by combining elements of previous experiments) will be a particular goal of these efforts.

Building Blocks: Research in the 4-5 year time frame will focus on incorporating and sharing the new building blocks of experiments that were created of extended in the 3-year frame, and continue to be developed to be used with open interfaces (see above). Incorporating the new building blocks consists largely of efforts in integration and extension (see below). Additional work, however, will focus on the sharing of old and new types of experiment building blocks. Increased sharing and re-use will depend in part on new (or extended from the 3-year time frame) types of metadata to describe experiments (or elements or components of experiments), and ontology to link them into a coherent whole. Ontologies will in turn enable the development of new sharing capabilities, via new approaches to cataloguing and searching knowledge bases about experiments and experiment operation.

Integration: Integration will continue to be an important area of work in the 4-5 year time frame, needed to demonstrate how to weave together new or extended building blocks, use new or extended open interfaces, and integrate their use into workflows of experimenter activity.

Long Term – 10 Years

The 5 to 10 year time frame will include (but certainly not be limited to) research work that tackles currently difficult problems, especially those that require the extended infrastructure and methodology resulting from the initial 5 years. Among those research areas that can be imagined now are:

Workflow: Advanced interfaces for experimenter interaction, including approaches to capturing experimenter intent.

Interfaces: Support for extensibility of facilities' capabilities to meet the need of research of domains that are beyond the horizon; together with natural evolution in the tools, techniques, models, and processes of experimentation that developed in the 1 to 5 year time frame.

Building Blocks: New building blocks needed for over-the-horizon domains; as well as advanced techniques for cataloging of experimentation approaches, metadata definitions, ontologies, and knowledge bases, where indexing and searching can incorporate high-level semantics and use consistency mechanisms across a large number of facilities the interoperate or enable sharing between their experimenters.

Integration: Automatic integration, interface abstractions that enable plug and play interoperation between many facilities.

4.4 EXPERIMENT DESIGN AND INSTANTIATION

Initial Description of the Capability

Research in cybersecurity requires sound science that builds upon valid, repeatable experiments. The community needs tools that will help researchers rapidly craft meaningful, validated experiment designs and create test environments that are based on real world models. The problem space may be divided into five subcategories:

1) Design tools, specifications, ontologies, and compiler
2) Reusable designs for science-based hypothesis testing
3) Automated discovery of local and distributed resources
4) Dynamic instantiation of domain-specific test apparatus
5) Validation of instantiated test environments and apparatus

This set of capabilities is roughly analogous to the code editor, libraries, compiler, and linker portions of an Integrated Development Environment (IDE) for software.

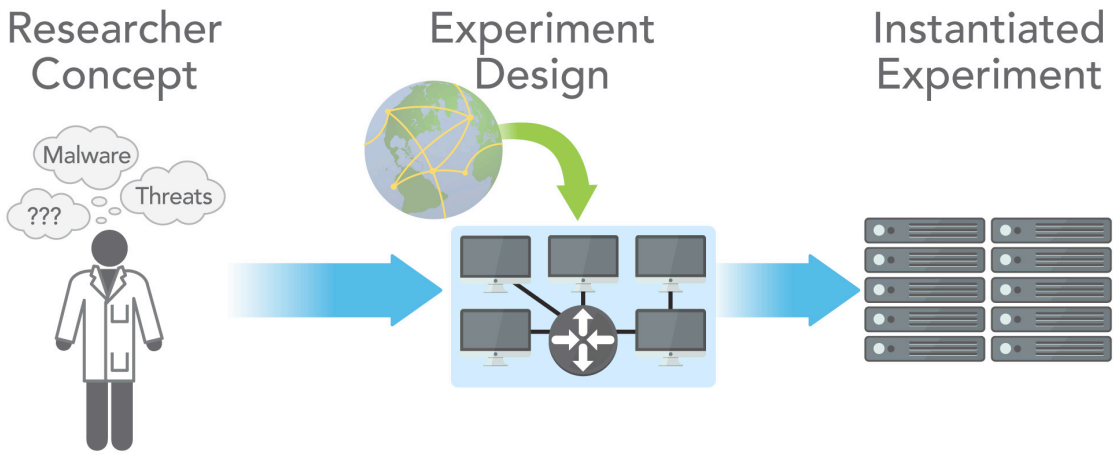


Figure 7. Design tools help researchers rapidly craft meaningful experiments and create test environments based on real world models.

Where do we want to be in the future? What will solutions look like? What will the solutions enable us to do?

Researchers across all domains that rely on computational systems will rapidly design meaningful experiments that reflect the real world by reusing and extending other experiment designs and design components. Experiment designs may focus on a single domain or span across multiple interconnected domains. Experiment designs will be automatically validated and design error warnings provided. Validated designs will be automatically realized in test environments, which will then be externally validated against the design.

A robust core set of design methodologies, tools and languages, specifications, and ontologies are needed. These capabilities will enable an experiment designer to specify the test environment, including domain-specific features, actors and environment interfaces,

events (to include attacks, faults and other real world events that might change the system operation and user decisions), and instrumentation and data to be collected (see Section 4.7). Orchestration design tools will enable researchers to create rich experimental scenarios based on models of real world activity (see Section 4.2). These scenarios will include alternate execution paths based on success of prior steps, time, and other environmental triggers and will be used in experiment execution (see Section 4.6).

A wide range of community component libraries is necessary to facilitate reuse and sharing. These libraries will contain domain-neutral and domain-specific components, expressed using rich a rich ontology. The libraries will include:

1) Reusable experiment designs and proven design patterns that may be used as guides for developing new experiment designs
2) Physical components (e.g., network topologies, standard servers, and specialized devices)
3) Actors (e.g., real or simulated system users and attackers) and actions (e.g., user behaviors, attacks)
4) Calibrated test data sets
5) Standard and specialized instrumentation, (e.g., network packet capture and log parsers)

An experiment designer will search the library, choosing which elements she wishes, and include those elements in her experiment design. Experiment designers will create new elements and optionally contribute them to the global library for use by others. Experiment designs may also be shared with others in a global library.

A compiler-like and lint-like tool suite is needed to provide syntax checking and to validate design specifications, ensuring all components are used properly and that the overall design is complete.

Automated discovery of local and distributed resources are needed to instantiate an experiment design within a test apparatus. Researchers need a rich search capability for available facility and experiment resources. Also needed is a means to dynamically instantiate a domain-specific test environment from the specified experiment elements or building blocks. This includes automated dynamic integration of widely distributed elements under different administrative control for experimental use.

Finally, an improperly configured test apparatus can invalidate results, and sometimes do so without the researcher's knowledge. Unknown interactions or mismatches between the experiment and the test apparatus also lead to invalid results. Validation of instantiated test environments is critical to increase the probability of experiment validity.

Table 8. Summary of Current State, Vision, and Needed Research for Experiment Design and Instantiation.

Capability	Current State	Vision	Needed Research
Design tools, specifications, ontology, compiler	Limited approaches in GENI, NCR, MIT-LL	Rapid design of real world experiments across multiple domains Reuse of experiment designs, components Designs automatically validated	Leverage work in ontologies and patterns from SW engineering Create pattern recognizers and sharing capabilities for reuse Techniques to analyze experiments and hypotheses to identify potential issues
Reusable designs for science-based hypothesis testing	Non-existent	Common sound design patterns are captured and reused in experiments within a domain	What is meant by reusable, and repeatable Capturing hypothesis, cataloging salient attributes
Automated discovery of local/distributed resources	Almost none existent	Rich search of facility and experiment resources; automated search for resources based on experiment specifications	Cataloging, metadata, distributed libraries, sharing, automation
Dynamic instantiation of building blocks for domain-specific test apparatus	Moderate ability to automatically instantiate standard IT network/service designs exists for both physical and virtual test environments Almost none existent for other domains	Dynamic realization of test environment designs for all domains Automated inclusion of widely distributed building blocks under different administrative control	Valid ways to combine device simulations with real devices and networks to emulate the real world Dynamic instantiation, composition via common interfaces (dependent on building block activity, see Section 4.3)
Validation of instantiated test environments and apparatus	No validation against real world in test facilities, except for self-validation by experimenters Other kinds of needed validation is limited	Automated correctness and real world applicability validation of test environment and apparatus instantiation, accounting for the degree of fidelity	Facility validation metrics; theoretical work in definition of non-interference for experimentation; models of correctness, validity Validation of test environments against models of the real world

Where are we today? What are the current technology and research? What are the important technology gaps and research problems?

Tools to support experiment design today are limited in the IT domain and mostly non-existent in other domains. Graphical network design tools for IT research are available but many researchers still manually construct network topologies in network topology design language files (e.g., ns-2 [34]). Network design tools for experimentation are mostly non-existent in other domains. Other experiment design aspects are unsupported -- for example, combining existing experiment designs and components to compose a new design. Usability studies are needed to identify design workflows and procedures and to determine how design support tool user interfaces should function in different domains.

All experiment design elements must be validated to ensure confounds are not inadvertently introduced into the experiment. Additionally, the experiment analysis method must be validated to ensure the soundness and conclusiveness of results. Common errors, such as failure to use the right statistical power, can result in inconclusive results. Validation of experiment designs today is performed manually by researchers and is error prone, resulting in wasted time and effort. In some cases, researchers may be unaware of such errors and incorrectly assume their experiment results are conclusive. Such errors can, if undetected, lead to cascading failures in future efforts that rely on such questionable results. Hence, validation is a key research issue that must be addressed in cyber experimentation of the future.

Ontologies must be extended to enable rich hypothesis specification and experiment design descriptions, such that automated validation can be performed, much like program verification and the syntax checking of compilers. Also necessary is an understanding of the salient features of an experiment design that must be captured and the pre and post conditions (e.g. constraints and invariants) necessary to judge correctness.

Pattern recognition analysis, along with other relevant techniques, should be explored for use in automatically deducing reusable design patterns from existing experiment definitions. Searchable libraries of known, reusable sound experiment design patterns should be created for sharing and reuse. This requires fundamental research to understand what is meant by reusable and repeatable.

Finally, a key missing item in research today is vetted, benchmarked, openly available data sets. The need for such data transcends merely the data requirements to conduct individual research efforts. The lack of these data sets directly impacts repeatability, which is a key cornerstone of good science. If benchmarked data is available for wide use, researchers will be better equipped to validate each other's work. In addition, the availability will enable true apples-to-apples comparisons of different approaches and algorithms for specific problems. Very little benchmarked data exists today, and what does is not openly accessible.

Table 9. Research Milestones for Experiment Design and Instantiation.

Capability	Near-term	Mid-term	Long-term
Design tools, specifications, ontology, compiler	Leverage existing ontologies Expand existing experiment design tools to support ontology and design patterns User studies for multi-domain, scalable design interface	Ontology domain extensions Fundamental research into compilable design languages Capabilities for reuse building on ontologies and patterns	Increased automation, real-time feed back to-from ontologies Semantically rich hypothesis and experiment specification Automated hypothesis and design analysis tools that flag issues that may create confounds or otherwise invalidate experimental results
Reusable designs for science-based hypothesis testing (A Top 5 Recommendation)	Specification standards for design patterns from SW engineering Creation of pattern recognizers and simple sharing mechanisms for reuse of designs and design patterns	Definitions of ranges of repeatable and reusable, formal specification of degrees of tolerance	Automated design pattern capture and increased specification for sharing Capturing and reasoning about hypothesis, cataloging salient attributes, creation of pre/post conditions for experiment correctness
Automated discovery of local/distributed resources	Offline, out of band search of central repository	Distributed resources repositories and descriptors, distributed search, manually guided	Increased automation, distributed search, discovery via automated meta information and locators
Dynamic instantiation of building blocks for domain-specific test apparatus	Transition automated test environment instantiation capability from government, university, and industry labs Building blocks integrated into experiment definitions, static and manual	Automated test environment instantiation for other domains (e.g., transportation, electric power) Dynamic allocation of remote/distributed resources as part of experimentation	Automatic inclusion of building blocks from template into running experiments for both local and remote resources

Capability	Near-term	Mid-term	Long-term
Validation of instantiated test environments and apparatus	Experiment apparatus diagnostic test procedures tuned for specific environments, verify non-interference by observation Assisted validation of instantiated test environments against real world models of the IT domain	Common standards of validation, validation across multiple disparate infrastructures, definition of metrics for instrumentation for self-check Fully automated validation of instantiated test environments against simple real world models in multiple domains	Theoretical work in non-interference definition for experimentation; models of correctness and validity with respect to fidelity Fully automated validation of instantiated test environments against complex, large-scale real world models in multiple domains

What will it take to get us there? How can we divide the problem space? What can be done in the short (3 years), mid (5 years), and long term (10 years)? What resources are needed?

Near term – 3 Years

Fundamental research breakthroughs are needed to provide a total experiment design tool suite. This research will take a number of years to complete. In the near term, a partial first-order design chain capability, consisting of a semantically rich design language and visual design editor, should be provided. A base ontology should be developed using the IT domain ontological work from the GENI [3]and NCR [4][5] communities. This ontology should include common terms from all domains of interest. IT domain specific extensions should be separated into an add-on module to be replaced or augmented as needed by other domain specific term extensions.

Existing experiment design tools (e.g., USC-ISI's MAGI [35]) should be modified to use the base ontology and domain extensions. User studies should be conducted to understand the required characteristics and properties of a user interface that sufficiently and efficiently supports the design of multi-domain, scalable experiments. The chosen design tools should be used to test user interface concepts and iteratively roll out incremental improvements.

In the area of reusable experiment designs, both full designs and design templates are needed. The concept of software engineering design patterns should be borrowed to create a base set of good, reusable experiment design patterns or templates for the IT domain. Pattern recognizers should be developed so that design patterns may be automatically extracted from "gold standard" experiments. An open, easy to use repository for sharing experiment designs and design patterns should be created and deployed. Design lookup capability should be integrated directly into experiment design tools.

In order to achieve the vision of having the ability to dynamically create larger testbeds by combining resources from multiple labs, we need the ability to dynamically discover

available resources that may be used in an experiment. A means to adequately describe what assets exist, when they are available, and who may use them is needed; this is another place where the GENI ontological work may be leveraged. A repository of available resources should be built and it should be manually searchable by researchers. Researchers can then use the knowledge gained from the search in designing experiments.

Much work has gone into technology to automatically instantiate testbeds from testbed specifications. Utah's Emulab [33] is one such effort, which manages and instantiates IT domain testbeds from physical machines. MIT-LL also has technology in this area for virtualized testbeds. These and other similar technologies provide a good foundation and lessons learned that should be transitioned to public use immediately. IT domain experiment building block specifications should be available for static and manual inclusion in experiment definitions.

Validation of the instantiated test apparatus and environment is critical to scientifically sound experimentation. In the near term, diagnostic test procedures tuned for specific environment should be developed and used. Recommendations for test procedures to validate non-interference of the IT domain test apparatus should be created and disseminated, and a validation checking assistant should be created to validate IT domain test environments against real world models.

Mid Term – 5 Years

In the mid-term, ontological domain extensions should be created for non-IT domains, such as transportation, medical, and electric power. Fundamental research should be conducted to develop compilable experiment design languages and language extensions for non-IT domains.

Existing ontologies and patterns should be leveraged and extended to enable reuse across all experiment design components.

Effort should be applied to understand what it means for different kinds of experiments to be "repeatable" and what information must captured so that peers can repeat and validate experiments. Ranges of "repeatable" and "reusable" must be defined, along with a formal specification for degrees of tolerance.

The repository of available resources should be extended to support manually guided, distributed searching to locate and include remote resources in a local experiment. In addition to equipment and related resources, the ability to easily share benchmarked data sets and incorporate into experiment designs is needed. Section 4.7, Instrumentation, discusses the need to collect and properly benchmark data. Sharing mechanisms must consider the policy necessary for data sharing. The ontology must be rich enough to capture and convey a data set's pedigree, which includes the method by which it was collected. Finally, the concern of personally identifiable information must be addressed. Numerous efforts have been underway researching means to sanitize data while preserving the salient features needed for research. This work should be considered and adopted where appropriate.

The IT domain automated test environment instantiation should be transitioned to other domains of interest, such as transportation and electric power. The automated instantiation

capability should be extended to support the dynamic allocation of distributed, remote resources in support of an experiment.

Significant effort must be applied to validation in this period. Common standards for validation should be developed. Special attention should be paid to validating across disparate infrastructures and domains. Metrics for instrumentation self-checks should be developed. By the mid term, the validation of an instantiated test environment against simple models of real worlds should be automated for experimentation in multiple domains.

Long Term – 10 Years

Automation should continue to increase to include real-time feedback at the 10-year mark. Automated pattern extrapolation should be feasible at this point, coupled with more a more robust and mature specification language for sharing designs and components. Researchers should be able to fully express hypotheses and experiment designs in a semantically rich way.

The ability should be developed to automatically capture and catalogue the salient features of an experiment, based on what is necessary for repeatability. This capturing should include the pre and post conditions necessary for experiment correctness. Tools should be developed to automatically analyze the hypotheses and designs and provide immediate feedback to the researcher on any confounds or other issues that might invalidate or otherwise result in questionable results.

There should also be an increase in automated support for researchers in searching for distributed, usable experiment resources. Both local and remotely discovered resources should be automatically included into running experiments.

Theoretical work in the definition of "non-interference" should be maturing by the 10-year mark. Models of correctness and validity with respect to fidelity should be mature and ready for use. Test environment validation against complex, large-scale real world models that span multiple domains should be fully automated at this point.

4.5 INTERCONNECTED RESEARCH INFRASTRUCTURE

Initial Description of the Capability

Future experimental research infrastructure must have a much more flexible and automated capability for interconnection than current technology provides. Top-level goals, properties, or requirements for interconnection include:

1) Automation of interconnection that enables transparency of sharing resources
2) Interconnection that can be dynamic or on-demand, but controlled to conform to specific models of interconnection and specific policies of resource sharing
3) Support for a variety of types of experiment using a variety of types of interconnected resources including real, virtual, emulated, and simulated resources.

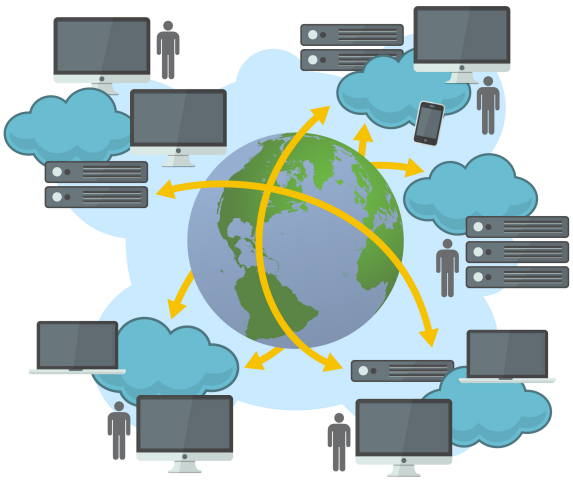


Figure 8. Flexible and automated capability for connecting research infrastructure.

In each of these three areas, broadly applicable but focused research will be needed to bridge the gap between the current various ad hoc modes of sharing, in order to create a common but flexible framework for interconnection and sharing between many testbeds and cyber-research facilities. As the sophistication and reach of basic testbed technology enables the creation of more testbeds – and more research infrastructure based on them – such a framework will be required to replace the current ad hoc pairwise interconnection approach, which will not scale from 10's to 100's of testbeds and from 100's to 10,000's of possible inter-connections, over the next 5 years.

Where do we want to be in the future? What will solutions look like? What will the solutions enable us to do?

For future interconnection of research infrastructure, the general principle is the same as at present: the unit of control of global cyber-research infrastructure will be a single instance of a testbed; multiple testbeds will have a common capability to interconnect with one another. Such testbed-level interconnection will support sharing of resources so that a researcher, whose research infrastructure relies mainly on one testbed, will be to use resource from that facility and others that allow such sharing.

However, future sharing must be based on not only on interconnection and sharing of computing and network level resources of a testbed, but also on inter-operation between cyber-research facilities, i.e., facilities that operate a testbed as the underlying basis for experimentation, and also encapsulate resources using advanced capabilities for creation and management of repeatable experiments.

One difference is a shift from per-experiment sharing to true inter-testbed sharing. Currently, an experimenter in one testbed can use sharing techniques to link in resources from another testbed, so that the experiment uses resources from both. This approach is important for variety, scale, and realism of wide-area network usage, yet the responsibility for the sharing set-up rests largely or entirely with the "power user" experimenter who can use current sharing techniques. While these power-user techniques will no doubt continue, higher-level and more approachable techniques will be needed for future experimentation, where a priori establishment of sharing is not required of each experimenter.

By contrast, future sharing should start at the level of inter-connections between testbeds, and sharing of resources between testbeds, rather than sharing explicitly defined at the level of an individual experiment. True inter-testbed sharing rest on a principle that is simple to state: the resource management infrastructure of research facility A should enable its operators to designate some resources as accessible by users of other facilities; and the resource management infrastructure of research facility B should enable operators or users of B to use those resources made available by A. Users of B should then be able to create experiments with resources from A and B.

Combining these two future advances – explicit sharing between testbeds and facilities as the pre-condition (rather than as at present the result) for federated experiments; and sharing of resources defined by rich a flexible object system of sharable resources of many kinds including compositions or collections of abstract resources – will require advances in combining two other areas: resource discovery and management, and access control.

Resource discovery is a blanket term used in broad areas of computing research, and covers a wide range of practices and technology including directories, publish/subscribe information distribution, tagging, search, and more that at present is largely disconnected from the infrastructure of cyber-experimentation. Adoption of the results of this parallel research will be required for its own sake, but it will also be combined with results of research on access control for federated access. Because federated access occurs without centralized structures for identification and policy, future interconnection and sharing will be based on a flexible ontology of actors, objects, attributes of each, allowed or disallowed access modes, and a trust management system for managing assertions about actors or objects.

As a result of these advances, experimenters will have a view of available federated resources that is on a par with their view of resources available in the (or each) individual facilities to which they have direct access. Experiment management capabilities will encompass all of these resources and the access to which the experimenter has been granted. An experimenter may specifically request particular federated resources, or be completely unaware of the extent to which experiment realization uses federated resources – and several other points on the continuum between full transparency and complete abstraction.

Table 10. Summary of Current State, Vision, and Needed Research for Interconnected Research Infrastructure.

Capability	Current State	Vision	Needed Research
Automated, transparent federation to interconnect resources	Experiment federation in DETER and GENI, requires a priori establishment	Fully automated using discovery approach described above, connection of wide range of aspects (experiments, testbed, specialized components)	Flexible interconnection semantics, auto negotiation, access control specification and mechanism
Dynamic and on demand, with sharing models	Little to no dynamic and on demand sharing Mostly closed models. DRAGON provides some dynamic short lived interconnection semantics	Highly dynamic interconnection Experimental Models that span multiple aspects	Dynamic establishment of interconnection, short lived sharing of resources, resource allocation at multiple abstractions Modeling of complex shared attributes
Support integrated experiments that include real, emulated (virtual), and simulations	Most support fixed level of abstraction with some mix of emulated, simulated and virtual (cf DETER, Emulab)	Support wide range experimental infrastructure bare metal, virtual machines, emulated CPUs, networks, simulations, in any combination, to create the appropriate phenomena required for validity in that part of an experiment	Approaches to mixed abstraction in specification and experimental realization, methods for assessing fidelity tradeoffs and validity (see Section 4.2)

Where are we today? What are the current technology and research? What are the important technology gaps and research problems?

Presently, there are several organizations, testbeds, or facilities that could be considered as operating or consisting of cyber-research infrastructure, but with a wide variety of missions or characteristics. Of these, relatively few are capable of or operated to include inter-connection. Of the few that do support inter-connection, the interconnection is pair-wise and driven by implicit policies or characteristics of each testbed.

Current technology and research includes two inter-related areas relevant to inter-connection and sharing. First, at a lower level of abstraction, the term "testbed" is often used to describe a facility that presents to its users a set of computing and networking resources that the user can assemble into a test network within which the user can build systems to be tested, or to be observed to learn about behavior or properties of the software or systems be tested. Across the variety of existing testbeds and facilities, current research includes work that continues several years of development of testbed technology, that is, increasingly sophisticated technology for managing and using a testbed's resources.

At the level of testbed technology, interconnection between testbeds (whether traditional testbeds, or the testbed layer of a cyber-research facility) interconnection consists of ad hoc "network plumbing" that is needed as a practical matter to support communication between testbeds. The ad hoc inter-connections are created mainly to meet specific needs of resource sharing at the level of an individual experiment that is based in one testbed and requires resources from another

At a slightly higher level of abstraction, access control to specific sharable resources is part of current research (on the technology base for cyber-research facilities) and practice (the use of cyber-research facilities). However, there is little in the way of organized infrastructure for global resource discovery and usage described above. By contrast to interconnection per se, the mechanisms of resource sharing are oriented to enabling an experimenter to explicitly set up resource sharing at the per-experiment level of abstraction. It is up to the experimenter to identify resources, request interconnection connectivity, request access to specific resources, and build an experimental apparatus that spans multiple facilities or testbeds, that is, a federated experiment.

Current research consists mainly of work on this type of resource federation, and extension of its enabling technology, at least in the context of cyber-experimentation infrastructure technology and usage. In the broad research context, there is a wide but unfocused body of work on full or semi automated resource discovery (both enterprise level and wide area), resource metadata schemas, and full or semi automated tagging, and similar topics at a very high level of abstraction that is not specific to research testbeds or facilities.

For each of the three major capabilities or goals described above, there is little or no directly targeted research, and – again with the notable exception of resource-level federation – each is essentially an open research area rather than a research with specific technology gaps.

Table 11. Research Milestones for Interconnected Research Infrastructure.

Capability	Near-term	Mid-term	Long-term
Automated, transparent federation to interconnect resources	Common framework for basic interconnection Resource publication and discovery for current/common resource types	Inter-connection authorization policies, tools to manage policies, mechanisms to implement inter-connection Policies limited to current/common resource types	Interconnection and shared resource access fully supported for all types of resources and many modes of access to them Automated negotiation of access and enforcement of access controls

Capability	Near-term	Mid-term	Long-term
Dynamic and on demand, with sharing models	Dynamic resource discovery and automated request for sharing, but human operators involved in setting up interconnections, and in authorizing access to shared resources	Initial inter-connection and/or federation driven by policy model and tools, but authorized with operator involvement After initial steps, on-demand access to resources shared within an inter-connection	Multiple models for sharing, fully supported in policy and tools Fully dynamic inter-connection, fully dynamic resource sharing in an inter-connection, for parties that fit an existing sharing model
Support integrated experiments that include real, emulated (virtual), and simulations	Integrated experiments with current/common resource types of real and virtual	Sharable resource types extended to one or more common model of simulation, emulation	Sharable resources include large compositions of resources

What will it take to get us there? How can we divide the problem space? What can be done in the short (3 years), mid (5 years), and long term (10 years)? What resources are needed?

Near term – 3 Years

In a 3-year time frame, research can be partitioned in separable areas in which partial results can be demonstrated, though continuing research will be required.

Interconnection techniques, policy, and enforcement can be a relatively near term area of research, focused on policy enforcement using existing mechanisms for network-level connection between the networks of two facilities that will share resources. At root, the mechanisms are likely to be similar to virtual private networks, with dynamic setup/teardown on-demand by the requirements of higher-level sharing over the network connection.

However, elements of resource discovery trust management, access control policy definition, and policy enforcement will all play a role. For example, one facility may advertise that it is willing to set up an interconnection on demand, with specifics of IP addressing, IPsec certificates, port usage, etc., as needed by another facility to set up its end of an interconnection. However, one facility also would likely have a policy on what kinds limited access would be enabled by interconnection – or multiple policies for multiple types of interconnection or other facility. Although these policies may well be expressed at typical low level of network address space, ports, and protocols, nevertheless there must be at least a basic policy management infrastructure for binding policies to dynamically constituted interconnections.

In a 3-year time frame, the basics of dynamic interconnection mechanisms, and the administration of them, can be defined, implemented, and trialed out in experimental interconnections, to determine the types of advanced capability that might be the focus in a 3 to 5 year time frame. Also, fully dynamic on-demand interconnection is not strictly

necessary; to have a lower bar for participation in pilot exercises, both initiation of interconnection requests, and complete of interconnections at the network level, can and likely would be mediated by human administrators.

Resource sharing ontology can be the focus of efforts in the early part of a 3-year time frame. The ontology should cover a significant subset of an overall ontology of resources types and definitions, resource attributes, modes of sharing and corresponding access modes, access controls, actor types and definitions, identification and authentication infrastructure, authenticated assertions of attributes of actors and resources, and entities that make such assertions.

Initial development of a common framework for interconnection between facilities will be enabled by the combination of the partial ontology and the basic inter-facility sharing communication and control mechanisms. The initial framework can be limited to common-denominator resource types (such as real or virtual hosts and real or emulated network connections between them) as well as basic use of federated external resources, e.g. an experimenter's local workstation integrated into an experiment.

In the 3-year time frame, federated experimentation can continue to work within the current types of basic sharable resources; extending the types of resources shared is less important in the near term, than building the infrastructure required for this shift.

Domain-specific resources should be a focus for short-term research by cyber-experimenters in specific domains of cyber-security research. Greater variety and availability of type of resources will be a significant benefit of the results of work on interconnection and resource sharing. However, experimenters within specific domains need to develop resources to share.

Particularly as the "Internet of things" expands the scope of connected systems (some of which are mission- or safety critical, while others are potential threat vectors), it will be particularly important that a wide variety of devices be available for test and experimentation, so that cyber-security research develops in parallel with the expansion of assets to protect. Particularly in physical critical infrastructure system, it will be important to develop sharable resources of simulated components, in order to achieve federated experiment scale beyond the scope of available real instances of CI components.

Mid Term – 5 Years

In a 3-5 year time frame, research in each of the above areas can proceed to significant milestones. Interconnection techniques, policy, and enforcement as well as a common framework for interconnection in the mid term should focus on extending the basic policies of short-term research. A much more complete ontology of resources will be required. A rich set of policies will be needed so that the interconnection framework can accommodate a cyber-research facility whose operators wish to set up interconnection capability with a variety of other facilities, with a variety of different access control policies, both at the basic interconnection level and for specific types and instances of sharable resources.

Again, elements of resource discovery, trust management, access control policy definition, and policy enforcement will all play a role, but at a much greater level of flexibility and variety. As in the short-term pilot usage between facilities will be essential to prove the concepts of flexible, dynamic sharing. As in the near term, actual usage of the new sharing

techniques may still be mediated by human operators, but the principles of full automation and dynamism must be proved in the 3-5 year time frame.

Policy management tools will also need to be developed and subjected to proof-of-concept usage.

A broader range of sharable resources (defined by an extended ontology) is also a central feature of the 3-5 year time frame. All the types of resources – real, virtual, emulated, simulated, etc. – that are in common use the short-term, must be supported in the mid term.

Similarly, domain-specific resources, developed in the 3-year time frame, must also be accommodated in the expanded ontology, common framework for interconnections, sharing/access policies, and resource discovery infrastructure.

A variety of sharing mechanism may be needed for a variety of models for a testbed or facility acquiring external resources of various kinds, with varying degrees of trust and varying mechanisms to maintain testbed integrity while using unmanaged external resources. Even in today's world, much less a 3 year window, there are several types of external resources: entire unused physical hosts, VMs created as a sharable resource on a system being used for multiple purposes, VMs created in a cloud hosting environment; resource management models range from complete control by the testbed, to a SETI@home model [36].

Long Term – 10 Years

The research agenda for the 5-10 year time frame will be set by the results of research in the 3-5 year time, with specific areas of research to continue the first 5 years' trajectory toward the outcomes listed in Table 11.

Another major factor setting the 5-10 year agenda will be the evolution of computing and networking technology and practice in the first 5 years. Consider, for example, the commercial computing transformation in 2000 to 2005 engendered by commercial Virtual Machine Manager products such as VMware [37] and open-source products such as Xen [38]. Cyber-research facilities will need to track a similar 5-year evolution so that experimenters in facilities will have available to them a set of resources that includes then-current or emerging types of resource.

Inclusion of new resource types will have impact on all or most of the areas of research in the 3 and 5-year time frames. Consider for example software-defined networking (SDN) as one infrastructure-level evolution. If SDN becomes more common as a commercial computing practice, facilities may need include SDN technology as one aspect of experiments that researchers can construct, in order to accurately model real-world use of SDN. A corresponding extension of the interconnection ontology would ripple out through interconnection mechanisms, resource discovery, sharing policy, etc. As another example, consider the use of cloud-managed resources; in the 5-10 year time frame it may be possible to temporarily extend or even entirely constitute a testbed based on resources available from cloud service providers.

Whatever the shape of a 5-10 year research agenda may be for interconnection and sharing, the end target should be the culmination of the 3-5 year work, arriving at these important goals:

- A wide variety of resources, and of modes of access to them, will be supported for interconnection and sharing between facilities: common commercial resources, domain-specific resources, and new emerging types of resources; access to real resource, virtualized, emulated, simulated, etc. for all types of resources.
- Sharable resources include those that are composition of lower level resources into larger abstract collections that can be shared in toto.
- Sharing can be supported dynamically and on demand, with full automation of both negotiation of interconnection based on the policies of the interconnecting facilities, and enforcement of those policies on sharing that is enabled by interconnection.
- Use of a common framework for publishing interconnection availability, for acceptable methods of using then-available techniques for establishing interconnection, and for trust management between the interconnecting facilities.
- At the level resource sharing enabled by interconnection, use of variety of models for sharing policies, tools for management and administration of these policies, and integration of policy enforcement into fully automatic resource sharing.
- Support for the full range of resolution and visibility of sharing to experimenters, ranging from no awareness (shared resources are used as needed as part of a distributed on-demand resource pool) to complete specification by an experimenter of all and only shared resources to use.

4.6 EXPERIMENT EXECUTION AND MANAGEMENT

Initial Description of the Capability

Research in cybersecurity requires sound science that builds upon controlled, well-executed experiments. The community needs tools that will help researchers run and control experiments in test environments that are instrumented appropriately in a non-intrusive manner. The problem space may be divided into three subcategories:

1) Experiment orchestration
2) Visualization and interaction with experiment process
3) Experiment debugging with checkpoint and rollback
4) Experiment execution validation

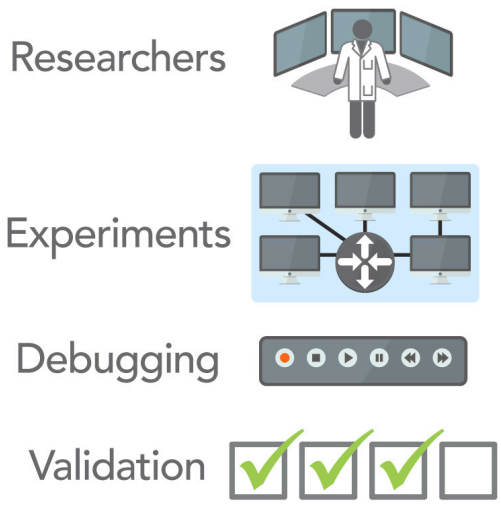


Figure 9. Tools to help researchers run and control experiments.

This set of capabilities is roughly analogous to code execution (and interaction, where necessary), dynamic runtime debugging and code stepping, and validating program behavioral correctness.

Where do we want to be in the future? What will solutions look like? What will the solutions enable us to do?

Researchers across all domains that rely on computational systems will orchestrate their test apparatuses to play out simple to highly complex, real world test scenarios. Experiment orchestration is the automation portion of running an experiment and includes both experiment execution and the management and control of test environment resources as execution occurs. Orchestrated scenarios, created by advanced experiment orchestration design tools (see Section 4.4), will include controlling or changing the behavior of standard IT systems and specialized domain devices. Scenarios will also include alternate execution paths based on success of prior steps, time, and other environmental triggers.

Researchers will be able to run both unattended experiments and interactive experiments in which they pause, analyze, and change the experiment flow. An interactive mode of experimentation is somewhat analogous to running an execution time debugger. This mode can be used not only for debugging purposes, but also for exploratory research and mixed modes of operation. Interactive mode will give researchers a visual window into the experiment as it runs and allow them to create checkpoints and pause execution when the checkpoints are reached. They may make environmental changes, if desired, and then resume execution or rollback to any checkpoint without invalidating the experimental results. Researchers will also be able to abort experiment execution, if desired. Intuitive visualizations will aide in human cognition in both modes of operation, helping researchers to quickly perceive the state of their experiment, as it changes over time.

The validity of experiments will be continuously analyzed at execution time. Researchers will be able to specify invariants at experiment design time (see Section 4.4) that are used by runtime validation tools to determine if the execution is valid. Test environment instrumentation (see Section 4.7) will provide the data necessary to verify the invariants are satisfied. This instrumentation will not interfere with experiment execution and will be sufficient to detect, record, and alert when unexpected conditions occur, when experiment validation failure is imminent, when an experiment fails, and when an experiment completes successfully. Researchers will use rapid diagnosis tools to aide in understanding failures so they may correct and quickly start again.

Table 12. Summary of Current State, Vision, and Needed Research for Experiment Execution and Management.

Capability	Current State	Vision	Needed Research
Experiment orchestration	Some capability for IT systems/networks to do simple orchestrations Little or none for other domains	Fully orchestrate test apparatus to execute and manage the steps of an experiment, inclusive of multiple paths of execution that depend on time, success of prior steps, changes in the environment, and other factors	Semantically rich orchestration language supporting full logic and conditional branching, with extensibility for domain specific experiments Domain specialized device control Pre and post conditions for experiment correctness validation Instrumentation for validation monitoring
Visualization and interaction with experiment process	Some capability for IT systems/networks, with advanced capability limited to government use Little or none for other domains	Ability to fully interact with experiments: pause, examine progress, make changes and resume, or restart Intuitive visualization provides full understanding of experiment execution state	Visualization approaches that convey experiment intent and help users quickly understand the impact of unexpected events on experiment integrity Resolve fundamental issue of how to suspend and resume system and network activity with no loss of integrity Experiment flow and dependency understanding and diagnosis techniques

Capability	Current State	Vision	Needed Research
Experiment debugging with checkpoint and rollback	Moderate checkpoint capability for virtualized hosts No checkpoint capability for real hardware hosts, although "deep freeze" technology could help Debugging is manual	Create and annotate multiple checkpoints to which one may choose to rollback Interact with experiments to debug at runtime	Resolve fundamental issues: how to capture and restore system and network state with no loss to experiment integrity What activity is sufficient to capture
Experiment execution validation	Manual process	Experiment execution is monitored in real-time, alerts of experiment validation status Automated support for rapid diagnosis of experiment execution failures	Fundamental research needed to understand what it means for an experiment to be valid or invalid Techniques to represent validity and test for it

Where are we today? What are the current technology and research? What are the important technology gaps and research problems?

The ability to orchestrate very simple experiment scenarios for the IT domain exists today. Outside of manually scripting scenarios, little to no scenario execution and management capability exists in other domains. Experiment scripting languages are needed, along with device abstractions that enable easy control of specialized domain devices as part of experiment execution.

Outside of using an off-the-shelf scripting language and building in interaction points, there is no general-purpose experiment interaction capability, and hence, no interactive debugging capability in any domain. Checkpointing is possible and supported for virtualized hosts, however no checkpointing is available for physical hosts and specialized devices within testbeds. It is feasible that "deep freeze" technology could be leveraged to provide an initial solution. Fundamental research will be required to understand techniques that may be used for checkpointing without invalidating experiments.

Some test apparatus instrumentation and test execution visualization capability exists for the IT domain. More advanced capability exists within U.S. Government and government contractor laboratories that could be used to provide an immediate capability, should the government choose to make the software available for general use. Little or no capability exists in other domains. The government use software could be extended to provide short-term capability to other domains.

Complete experiment validation, if done, is presently a manual process performed by researchers. No generally reusable experiment validation tools exist for use. Additionally, no rapid diagnosis tools exist for any domain. All diagnosis of experiment failure is performed manually by researchers and can be time consuming and incomplete.

What will it take to get us there? How can we divide the problem space? What can be done in the short (3 years), mid (5 years), and long term (10 years)? What resources are needed?

Near term – 3 Years

Experiment Orchestration: Building upon existing ontologies from projects such as GENI [3] and experiment orchestration tools such as the USC-ISI MAGI tool [35], create a semantically rich orchestration language and execution capability for the IT systems and networks domain. The capability should support the creation of orchestrations that use full logic with conditional branches. Include support for smart grid research using manual configurations and integrations of various smart grid embedded devices.

Table 13. Research Milestones for Experiment Execution and Management.

Capability	Near-term	Mid-term	Long-term
Experiment orchestration	Semantically rich orchestration language and execution capability, supporting full logic and conditional branching for IT systems/ networks domain	Support for other domains using manual configuration and integration of specialized, embedded devices	Automated configuration and inclusion of specialized, embedded devices for experiment orchestration
Visualization and interaction with experiment process	Transition existing experiment visualization and control capabilities from government, university, and industry labs Support tools for rapid diagnosis of experiment execution failures	Visualization approaches that convey experiment intent and help users quickly understand if and how issues that arise affect experiment integrity Semi-automated assistance for rapid diagnosis of experiment execution failures	Suspend and resume system and network activity with no loss of experiment integrity Fully automated rapid diagnosis of experiment execution failures.
Experiment debugging with checkpoint and rollback	Automation and inclusion of virtualization based rollback capability in experiment process control	Capture sufficient experiment state to support diagnostics and rollback	Rollback experiment state with no loss to experiment integrity
Experiment execution validation	Monitoring and warnings for conditions that may invalidate and experiment and require further (manual) investigation	Understanding of and ability to represent validity of experiment	Ability to automatically verify experiment validity

Visualization and Interaction with Experiment Process: A number of visualization and control capabilities presently exist at MIT-LL [15] and other locations. Transition these capabilities from any closed areas to be openly available and mature them as needed for use. Create tools that will help users rapidly diagnose experiment execution failures, both during execution and in post analysis.

Experiment Debugging with Checkpoint and Rollback: Enable a first order experiment checkpoint and rollback capability as part of overall experiment process control. Automate the control of the snapshot/rollback capability of virtualization technologies for this purpose, and integrate into experiment controllers.

Experiment Execution Validation: Experiment validity is not the same as experiment execution failure. While the failure of the execution will invalidate an experiment, the execution could complete normally and yet the experiment is invalid. Tools should be developed to automatically monitor for and detect common conditions that may invalidate an experiment and require further (manual) investigation. Issues alerts to the user when such conditions arise.

Mid Term – 5 Years

Experiment Orchestration: Add support for manual configuration and integration of specialized, embedded devices for domains beyond traditional IT systems and networks and smart grid. Manually integrate these devices into the framework so that experiment orchestration components can automatically use the devices in the same ways they would be used operationally.

Visualization and Interaction with Experiment Process: Develop visualization approaches that convey experiment intent. This will require the use of a rich, expressive language. Develop technology that uses experiment intent to assess and alert users when issues arise that may affect experiment integrity. Extend and begin automation of tools that will help users rapidly diagnose experiment execution failures, both during execution and in post analysis.

Experiment Debugging with Checkpoint and Rollback: Support checkpoint creation and rollback for non-virtual systems at multiple points in the experiment process. Warn if experiment fidelity will be impacted. Capture sufficient experiment state in the checkpoints to support diagnostics.

Experiment Execution Validation: Provide capability to correctly represent the pre and post conditions that must be met for an experiment to, during the various stages of experiment execution, be fully valid. This capability will require a rich representation language with conditional logic.

Long Term – 10 Years

Experiment Orchestration: Add support for automated configuration and integration of specialized, embedded devices all domains. Develop the capability to automatically analyze embedded devices, add an appropriate control interface, and integrate into the framework so that experiment orchestration components can automatically use the devices in the same ways they would be used operationally.

Visualization and Interaction with Experiment Process: Develop the capability to suspend and resume system and network activity during an experiment, so that researches can stop execution and examine system state and other items of interest based on alerts of

potential issues. This capability to suspend/resume should not result in any loss of experiment integrity. Fully automate tools that rapidly diagnose experiment execution failures, both during execution and in post analysis.

Experiment Debugging with Checkpoint and Rollback: Capturing all necessary system and network state, create checkpoints of non-virtual systems on demand. Support rollback to checkpoint with no loss to experiment integrity.

Experiment Execution Validation: Building on the experiment validation language, develop the capability to automatically verify experiment validity.

4.7 INSTRUMENTATION AND EXPERIMENT ANALYSIS

Initial Description of the Capability

Research in cybersecurity requires the collection, ensured integrity, and analysis of experimental data. The problem space may be divided into four subcategories:

1) Instrumentation and data collectors
2) Transport and protection mechanisms
3) Data repositories
4) Data analysis

This set of capabilities is roughly analogous to inserting debugging code and conducting post analysis of debugging outputs and log files created during execution.

Figure 10. Instrumentation and experiment analysis.

Where do we want to be in the future? What will solutions look like? What will the solutions enable us to do?

Researchers across all domains that rely on computational systems will utilize ubiquitous, non-interfering instrumentation for physical and virtual devices, including specialized domain-specific devices. One possible way to provide non-interfering instrumentation is to build a test bus into all compute platforms that is used and controlled by the experiment controller. Regardless of the implementation(s), researchers will be able to automatically generate and deploy or otherwise engage embedded instrumentation. They will designate which data elements to collect and store and how to pre-process the data.

Data collection will not exceed capacity of the data storage devices or networks. Instead, smart collectors will be dynamically controlled during the course of experiment execution based on real-time analysis, leading to "as needed" data reduction at experiment runtime. For example, detection of a specific trigger will turn on/off or re-configure a collector to only store the data necessary for the time and duration necessary.

Experiment data integrity and confidentiality will be ensured. Data transport from collectors to the repository will be fully protected against tampering, leakage, and misuse. The data transport mechanism will be completely invisible to the experiment and will not interfere with or otherwise influence experimental results.

Data repositories will be both human and machine-readable and will be searchable by both. Researchers will perform experiment post analysis using large collections of shared, extensible tools that operate on data in the local experiment repository. They will capture experiment results and extract lessons learned and share these so that other researchers may benefit from and build upon their work. Results and lessons learned will be captured using a rich, searchable ontology and be stored in an accessible community research knowledge store.

Experiment scenario actors, to include malware, will be automatically instrumented to produce needed ground truth information in real-time. Examples of information that might be captured include: what a simulated human or piece of malware did on a host and at what time, and how a system behaved when exploited. This information, along with any relevant system or software log files, will be collected and stored in the data repository for post analysis. Researchers will extend shared tools to rapidly recreate the integrated experiment ground truth for use in analysis of various approaches to research problems.

Where are we today? What are the current technology and research? What are the important technology gaps and research problems?

Experiment instrumentation consists of a few standard raw data capture tools (e.g., pcap [39], syslog [40], the Windows event log [41]) and hand crafted tools to post process existing data sources. There is no automated instrumentation capability for systems. Data collectors are generally static, unless specifically hand crafted to act in a more dynamic fashion.

Researchers use both in and out-of-band networks to control experiments and collect data. These approaches are not transparent to the experiment. Both include agents or other software on host sensors to collect host specific data. The in-band approach uses the same network as experiment traffic, so the collected data and experiment controls are present on the network and can be seen by and interfere with the behavior of hosts, actors, and malware. The out-of-band approach uses a segregated network for collected data and orchestration controls, so these are not seen. However, hosts in the experiment have an additional active network interface which can be seen by and interfere with activity (both benign and malicious) on the experiment hosts. In both cases, there are no enforced protections, so malware used in an experiment can interfere with and change control signals sent and data collected. Transparent, protective ways to control and collect data from both physical and virtual hosts are needed.

Lots of data repositories exist today, but many are unable to deal with the “big data” problem and do not support semantically rich searches on the data contained therein. There is no community-wide repository of shared experimental results, and researchers often are unaware of what other researchers are doing.

Generally, experiment analysis tools are custom built on an experiment-by-experiment basis. There are a number of basic server log and pcap analysis tools available as open source that researchers may use. Some cyber attack forensic analysis tools may be useful. Most existing tools are focused on system and network administrator tasks with a lessor amount supporting cybersecurity analysts. These tools could provide a core upon which to build, in some cases.

Table 14. Summary of Current State, Vision, and Needed Research for Instrumentation and Experiment Analysis.

Capability	Current State	Vision	Needed Research
Instrumentation and data collectors	Some standard (e.g., pcap, syslog), mostly hand crafted, rudimentary, low level No instrumentation automation or dynamic tasking of instrumentation	Ubiquity of non-interfering instrumentation for virtual, physical, static, dynamic, and real-time systems, that allows full or selective collection of data Automated instrumentation of specialized devices	Instrumentation techniques that are tamper proof and non-interfering Techniques to compensate for interference when such cannot be avoided Instrumentation specification language with techniques to automatically analyze and create instrumentation for specialized devices
Transport and protection mechanisms	In and out-of-band mechanisms are not transparent No enforced protections	Collected data transport mechanism(s) from sensors to repository that are fully protected against tampering, leakage, and misuse and do not interfere in experiment	Novel techniques for data transport that are tamper proof and non-interfering
Data repositories	Limited, centralized, and distributed storage facilities Many unable to deal with "big data" No data store exists for sharing results	Data repositories, which are fully searchable by both humans and machines	Semantically rich data description language Techniques for automated data tagging with full context, provenance, and auditability
Data analysis	Basic log analysis tools, forensics oriented with limited pattern analysis, focused on system and network administrators and cybersecurity analysts One-off tools are written on an experiment-by-experiment basis	Collections of pre-canned, yet extensible tools for post-experiment, multi-purpose analysis Real-time analysis and dynamic control of experiment instrumentation to perform "as needed" data reduction at experiment runtime	Tool designs for extensibility and multi-domain use Specification and validation of confidence bounds on experiments Automatic analysis and comparison of experimental results from multiple testbeds Correctly merging data from multiple sources, compensating for non-synchronized clocks Transformations to anonymize, normalize, and compress

Where are we today? What are the current technology and research? What are the important technology gaps and research problems?

A variety of tools presently exist for instrumenting experiments involving information systems, IT networking, and cloud computing. These include raw data capture tools (e.g., pcap [39], syslog [40], the Windows event log [41]), networking and telecommunications tools (SiLK [42], Argus [43], Bro [44], Snort [45]), BGP monitoring and analysis (Oregon's Route Views [46], RIPE NCC RIS [47], etc.), pcap analysis tools (TCPdump [48] and Wireshark[49]), and others. These tools provide a good foundation for collecting low-level data. Hand crafted tools are created to post process these data sources. There is no automated instrumentation capability for systems. Data collectors are generally static, unless specifically hand crafted to act in a more dynamic fashion.

Researchers use both in and out-of-band networks to control experiments and collect data. These approaches are not transparent to the experiment. Both include agents or other software on host sensors to collect host specific data. The in-band approach uses the same network as experiment traffic, so the collected data and experiment controls are present on the network and can be seen by and interfere with the behavior of hosts, actors, and malware. The out-of-band approach uses a segregated network for collected data and orchestration controls, so these are not seen. However, hosts in the experiment have an additional active network interface which can be seen by and interfere with activity (both benign and malicious) on the experiment hosts. In both cases, there are no enforced protections, so malware used in an experiment can interfere with and change control signals sent and data collected. Transparent, protective ways to control and collect data from both physical and virtual hosts are needed.

Lots of data repositories exist today, but many are unable to deal with the “big data” problem and do not support semantically rich searches on the data contained therein. There is no community-wide repository of shared experimental results, and researchers often are unaware of what other researchers are doing.

Generally, experiment analysis tools are custom built on an experiment-by-experiment basis. There are a number of basic server log and pcap analysis tools available as open source that researchers may use. Some cyber attack forensic analysis tools may be useful. Most existing tools are focused on system and network administrator tasks with a lessor amount supporting cybersecurity analysts. These tools could provide a core upon which to build, in some cases.

Table 15. Research Milestones for Instrumentation and Experiment Analysis.

Capability	Near-term	Mid-term	Long-term
Instrumentation and data collectors	Tamper-proof instrumentation Configurable instrumentation allowing for full or selective collection of data Non-interfering, instrumentation for virtual systems	Novel, non-interfering, instrumentation for physical, static, dynamic, and real-time systems Automated compensation for interference when such cannot be avoided Semi-automated creation of instrumentation for specialized devices	Automated instrumentation of specialized devices
Transport and protection mechanisms	Novel data transport mechanism(s) from sensors to repository that are fully protected against tampering, leakage, and misuse	Non-interfering transport mechanisms	
Data repositories	Human searchable data repositories, using "big data" analysis tools	Collected data automatically tagged with keywords, province, and audit information Data is machine searchable	Full experiment context of data is automatically inferred as data is collected, tagged to data, and is machine searchable
Data analysis	Integrated set of existing data collectors from multiple domains Ability to dynamically control instrumentation at experiment run time using signature-based detection of events *(e.g., something interesting is about to happen)* Ability to merge data from multiple sources without synchronizing clocks; assisted merging of collected data to construct ground truth in support of experiment measurements	Collections of extensible tools that provide semi-automated assistance for post-experiment analysis across multiple domains Specification and validation of confidence bounds on experiments *(when do results become questionable?)* Ability to anonymize experiment data Automated construction of ground truth in support of IT domain experiment measurements	Automated construction of ground truth in support of non-IT domain experiment measurements Collections of extensible tools that provide automated post-experiment analysis across multiple domains

What will it take to get us there? How can we divide the problem space? What can be done in the short (3 years), mid (5 years), and long term (10 years)? What resources are needed?

Near term – 3 Years

Instrumentation and Data Collectors: In the near term, focus should be applied to creating tamper-proof instrumentation. Tamper resistant technologies exist and should be leveraged to provide this capability.

It is presently infeasible to collect all data in many network-based experiments due to the vast amount of network traffic. As experiments increase in scale, our ability to collect the necessary data for post analysis will become unwieldy. In order to support experiments at scale, instrumentation should be configurable, allowing researchers to collect all or a reduced data set, as dictated by the experiment design.

Interference of data collectors can, at a minimum, call into question test results and may invalidate results. It is important to note that a non-interfering data collector in one experiment may actually interfere with other experiments: non-interference is context sensitive. A first order set of non-interfering instrumentation should be made available for virtual systems by leveraging the body of existing virtual machine instrumentation, such as that used to monitor malware. Honeypot technology may also be leveraged for this purpose.

Transport and Protection Mechanisms: In order to maintain the integrity of an experiment, the data transport mechanisms used to pull data from sensors to the repository must be protected against tampering, leakage, or misuse. While it may be unlikely that a researcher would purposely modify his data to change his results, accidental modification or modifying data without understanding the implications of such modifications can and does occur. Tamper resistant technologies are available on the open market and could be adopted to provide this capability early.

Data Repositories: Data repositories will necessarily grow as experiment scale increases. In order to aid human analysis of massive data, "big data" analysis tools and techniques should be utilized to reduce search times and increase positive results.

Data Analysis: One of the primary areas of focus needed to support cyber experimentation is on scientifically valid, multi-domain data analysis tools. Analysis tools should be developed to support integrated data from disparate data collectors spanning multiple domains. In addition, tools are needed to merge data from sources that may not have synchronized clocks, as all clocks in the real world are not synchronized and doing so may introduce unrealistic artificialities into an experiment. Finally, tools should be developed to assist researchers in merging collected data and constructing ground truth, which is necessary for experiment measurements.

Mid Term – 5 Years

Instrumentation and Data Collectors: A variety of novel, non-interfering instrumentation should be developed for physical, static, dynamic, and real-time systems. Hypervisors and other architectural constructs may be one approach to monitoring and collecting data off board for experimenting with technologies that do not use these constructs.

Transport and Protection Mechanisms: In addition to non-interfering instrumentation, transport mechanisms for collected data must also be non-interfering. Consider that data flowing across the experimental network during an experiment could interfere with an experiment involving a network packet analysis tool. Therefore, multiple means to transport data are required.

Data Repositories: As data size continues to grow, it will become more and more difficult for humans to search, find, and analyze data within large repositories. It will be necessary for data to be machine searchable. In addition, capturing and recording data province is critical so that if, downstream, it is determined that the data is flawed, research that utilized that data may be flagged as potentially invalid and in need of re-examination. When collected, data must be automatically tagged with metadata that includes its province and audit information. Automated assignment of keywords is also necessary to assist with automated machine searches.

Data Analysis: Tools are needed to support the sound use of a variety of analysis methods, including statistical methods. Collections of extensible tools that provide semi-automated assistance for post-experiment analysis across multiple domains should be developed. In addition, analysis tools should be developed to assist researchers in validating experiment execution. These tools should allow researchers to specify confidence bounds on experiments -- that is, to specify at what point experimental results become questionable, and at what point are results completely invalid.

The sharing of data is critical to experiment repeatability by peer researchers, which is a cornerstone of science. One of the main issues in sharing datasets outside a specific project is that data often contains personally identifiable information or PII. The creation of data anonymization tools that both protect individual privacy and ensure scientific validity must be a priority if we are to practice science in cybersecurity. A wide variety of data anonymization tools should be researched and available for use by this point.

Ground truth assistance tools should be extended to automatically construct ground truth for IT domain experiments.

Long Term – 10 Years

Instrumentation and Data Collectors: In the long term, a capability should be developed to automate the instrumentation of specialized devices, such as PDUs and embedded medical devices.

Data Repositories: In addition to automatically instrumenting special devices, the capability to automatically infer and tag the full experiment context to data should be developed. As with other metadata, this tagging should be machine searchable.

Data Analysis: Tools that construct ground truth for IT domain experiments should be extended to construct ground truth for other domains. Analysis tools should be extended to perform analysis across multi-domain experiments.

4.8 META-PROPERTIES

Initial Description of the Capability

In addition to the core capabilities described in the previous sections, there are several meta-properties of importance. Research infrastructure needs to be easily usable by both a wide range of experimenters and by the owner operators of the infrastructure. Then, in support of experiment validity and as described in Section 4.6, there need to be features of research facilities that provide confidentiality, availability and integrity of the experiment ecosystem; underlying security mechanisms are required, but also security policies and models, user interfaces to experimenters, and administration capabilities for the operators. Finally, there are a number of cultural and social changes along with community building that will facilitate future capabilities. These are captured as requirements for:

1) Usability (experiments, owner/operator)
2) Confidentiality, availability and integrity of experiment ecosystem
3) Social and cultural changes

Figure 11. Usability, security, and social and cultural changes.

Where do we want to be in the future? What will solutions look like? What will the solutions enable us to do?

Usability: Future cybersecurity research infrastructure will be usable by: a wide range of researchers and domain experts for each of several domains of research, not limited to traditional computer science researchers; a range of users, not limited to power users of

CEF infrastructure. CEF usability requirements involve: multiple areas of CEF usage; multiple areas of the "building blocks" for constructing and maintaining a CEF instance; multiple disciplines for designing and evaluating highly usable systems.

Given that the future research infrastructure is envisioned to be dynamic and spanning multiple usage models, it is important that usability work take into account not only usability of administration for a CEF infrastructure's owners and operators, but also usability of deployment of technology to create a CEF infrastructure. In the long term we can envision some degree of self-configuring or infrastructure-in-a-box capabilities to ease operational burdens, particularly for those prospective operators who are not from the traditional computer science background.

Given the breadth and variety both of CEF usage and usability issues, it is clear that there will be both variety and extensibility, if usability issues will have been adequately addressed. A variety of human interfaces to CEF infrastructure will exist, each oriented to a particular domain of research or type of role or activity enabled by the human interface. Further, human interface technology will be extensible and customizable, as the set of users and communities grows along with growth in the available infrastructure and its scope and scale.

Confidentiality, Availability and Integrity of Experiment Ecosystem: CEF infrastructure includes both a data management component – for data that includes experiment definitions and a variety of types of experimental data – and a subsystem for resource allocation and usage, that instantiates an experiment on the computing and networking resources managed by CEF infrastructure. Access control and availability are security functional issues for both.

Asset control will be relevant not only with CEF infrastructure, but also for sharing and interoperation between infrastructure instances. Such interoperation will be facilitated by extension of, and extended use of, current federation concepts such as federated identity and delegated access control.

In addition to asset access control, it will be important that CEF infrastructure includes both technical capabilities and best practices for establishing provenance of and ensuring the integrity of managed assets. Provenance will be increasingly important to support social and cultural changes that will be influenced by more effective and usable infrastructure for sharing. For one common example, in the case of an experimenter attempting to recreate another experimenter's experiment, provenance is critical for both experiment artifacts to be re-used, and experimental result datasets to be compared and analyzed.

Availability is another critical infrastructural meta-property of future infrastructure. As with the needed evolution of asset models and access control policy models for infrastructure asset access, there will also be an evolution of models for the use of resources for operating experiments. Facility owners and operators will have administration capabilities far exceeding basic resource reservation common today. Experimenters and users will have greater visibility on availability policies, and their effect on experiment operation.

Social and Cultural Changes: Along with usability-related and control-related meta-properties, it is clear that future CEF capabilities, and usage of them, will require cultural and social changes. Much usage of future CEF infrastructure will be characterized by the use

of evolvable frameworks that support evolution in experimental methods, and multiple models of collaboration, both within the user base of infrastructure and collaboration that spans multiple instances.

To enable these changes from the present situation, required activities include both research in several areas (CEF core capabilities, workflow and usability, frameworks and extensibility, sharing and collaboration, etc.), and concerted efforts by those researchers to stay engaged with the CEF experimenter/user communities that are the intended beneficiaries of that research in CEF capabilities and usage.

As these infrastructures and communities co-evolve, engagement with experimenter communities will need to socialize new CEF capabilities and experimentation methods via activities conducted specifically for user and community engagement. Such activities should bring together researchers and developers of CEF infrastructures, including exercises aimed at identifying commonalities and differences, and defining opportunities for collaboration between CEF experimenter/users and CEF infrastructure developers and researchers.

Where are we today? What are the current technology and research? What are the important technology gaps and research problems?

Usability: Usability is currently separately driven by the distinct needs of each of several very stove piped cybersecurity research communities grouped around the major existing research infrastructures and facilities, with only a small amount of cross fertilization. Within each infrastructure instance, usability is a function of user interfaces created for the users of that instance, but often oriented to the user needs perceived by technology developers, rather than the result of commercial best practices for product design activities such as workflow definition, user experience design, usability testing with real users and other stakeholders, etc. Even with such ad hoc approaches to user experience, existing facilities' usability can be "good enough" for specific communities, or for users with specific skill sets; but few would argue that any existing facility is highly usable across multiple skill levels and diverse user bases.

Table 16. Summary of Current State, Vision, and Needed Research for Meta-Properties.

Capability	Current State	Vision	Needed Research
Usability (experiments, owner/operator)	Varies by type of infrastructure and community, some highly usable by targeted communities, few to none usable across multiple skill levels and diverse user base, few to none with flexible user interface that can be evolved for increased usability	Adaptable usability tailored for a wide range of research communities, and range of user spanning wide base of domain experts to power users Facility owner operators from multiple domains use overarching frameworks for adaptable usability	Meta-research into effective and usable human interfaces to CEF capabilities Pedagogical tutorials, UIs specialized for self-teaching facilities, UIs specialized for power users or domain experts, adaptable set of defined common experimenter interfaces and workflows

Capability	Current State	Vision	Needed Research
Confidentiality, availability and integrity of experiment ecosystem	None to very limited, most existing research infrastructure developed as research Very little if any experimenter visibility on facility functionality for security meta-properties.	Best practices for establishing provenance of experiments and artifacts, controlled access to and integrity of raw data, experiment configurations, software, etc. Confidentiality of IP, availability and timeliness	New mechanisms for assuring integrity, and confidentiality of experiments and experiment data Tools and methods for automatic scheduling and resource management for increased availability
Social and cultural changes	Small community, somewhat in-grown, mostly network researchers building and using network testbeds, some stove pipe communities for CIP	Large multi discipline community of researchers operating across distributed and extensible research infrastructure working in an environment of sharing and libraries	Community efforts to socialize new methods and initiatives Bring together RI researchers and developers, conduct exercises aimed at identifying commonalities and differences and define opportunities for collaboration and integration

This open research problem should become more practical, with continued evolution of engineering work of CEF infrastructure, adopting current commercial practices such as web services interfaces that separate transactional core components from multiple web user interface components that have full flexibility to implement the user experience carried out via a web services interface.

The resulting meta-research into CEF human interfaces can then develop a variety of CEF human interfaces for differing purposes, rather than a one size fits all user interfaces. There may be human interfaces for differing purposes, e.g., those oriented to self-teaching of CEF usage and methodology, vs. those oriented to power users and/or specific domain experts.

Moreover, the wealth of human factors research and testing method can be applied to research experimentation that enables the applied engineering work is needed to achieve multiple goals: to close the gap between "usable by me" by-and-for-CS community, and broadly usable human interfaces for a range of experimenters; eliminate the need for one-size-fits-all user interfaces; create the building blocks of CEF human interfaces that enable adaptation and customization for usability factors specific to a specific CEF or research community.

Confidentiality, Availability and Integrity of Experiment Ecosystem: In current communities and facilities, there is very little focus on CEF characteristics related to the meta-properties of confidentiality, availability, and integrity of the experiment ecosystem. Obviously as security research infrastructure, individual operators take steps to ensure the

security of their systems, but providing widely available measures and assurances to the research community is rare. This stems largely from the fact that the majority of research infrastructure is developed under research programs that have evolved over time, rather than being developed to meet specific security functional requirements or specific operator capabilities for security.

It will be necessary to build in transparent and readily available measures of the security (confidentiality, availability and integrity) of the research infrastructure itself. Of course, creating mechanisms and transparent measures in this area will encounter long-standing problems around measurement and assurance as well as trade-offs between open and easily available research infrastructure and more closed systems.

Social and Cultural Changes: In terms of cultural and social community, currently there are a number of very stove piped cybersecurity research communities grouped around the major existing research infrastructures with only a small amount of cross fertilization. Furthermore the culture is still largely one of "build you own" and very little use of common or shared experiments or results. It will be necessary to simultaneously make progress on creating the frameworks, libraries and shared artifacts, while also changing community research methods.

Work on security meta-properties is relevant, particularly mechanisms for controlled sharing between infrastructure instances, to advance beyond today's "walled garden" model of some facilities. Usability of administration of sharing both within a instance and between instances is important as well. Both operators and experimenter/users need to see the usability, feasibility, and utility of CEF technology that both promotes cross-fertilization and enables controlled interactions.

Table 17. Research Milestones for Meta-Properties.

Capability	Near-term	Mid-term	Long-term
Usability (experiments, owner/operator)	Introduction of multiple points of entry to experimentation facility. Common operations and build environments, common language and terminology, extended operator community Separation of UI from core CEF capabilities	Meta-research into CEF workflow, methodology, human interfaces. Interface design, implementation, testing with commercial best practices Includes domain specific interfaces using the terminology, workflow and methods typically preferred within an established community-of-interest	Meta-experimentation with human interface building blocks, adaptation to domain specific needs, variation for levels of user sophistication Development of pedagogical human interfaces, training users for common experimenter interfaces, applications, processes, and procedures

Capability	Near-term	Mid-term	Long-term
Confidentiality, availability and integrity of experiment ecosystem	Protected experimenter space for isolation of sensitive data, increased facility controls for integrity Standard security evaluation of infrastructure	Base core experimental facilities of medium to high assurance systems, e.g., SE Linux, develop integrity checking processes, create scheduling tools, portals for availability	New mechanisms for assuring integrity, and confidentiality of experiments and experiment data New tools and methods for automatic scheduling and resource management for increased availability
Social and cultural changes (A Top 5 Recommendation)	Conduct workshops, disseminate to broader community, begin catalog and library collection in conjunction with frameworks (Section 4.3) and domains (Section 4.1)	Advertise new capabilities, require contributions back from users, maintain evolving libraries and search functions, require use of capabilities in certain funded efforts	Through education and socialization over time create environment that is so beneficial that people will seek it out and use it rather than inventing their own

What will it take to get us there? How can we divide the problem space? What can be done in the short (3 years), mid (5 years), and long term (10 years)? What resources are needed?

Near term – 3 Years

Usability: Essential innovations in future human interface software architecture are needed in the near term to enable innovative user interface modifications as the basis for research into human interfaces for several aspects of experimentation infrastructure:

- Workgroup collaboration and user management
- Asset management and access control
- Experiment lifecycle workflow
- Experiment operation and interaction
- Facility management and administration
- Policy management for security and availability
- Policy management for interoperation and sharing among infrastructure

and other user-facing or operator-facing innovations envisioned in this roadmap.

To enable the beginnings of policy modeling and user experience design, the basic innovations are the separation of user interfaces from core CEF capabilities described above, and the consequent ability to introduce multiple points of entry to CEF infrastructure – particularly new human interfaces (and new programming interfaces to support them) to be developed, tested, and investigated as part of meta-research on human interfaces for research infrastructure.

Further, to enable results from this work to be transferrable among different instances of experimentation infrastructure, it will be required to define and implement a set of

operations that can be a common functional subset supported by multiple instances; and a common terminology and ontology to support it. To be sure, it is up to an infrastructure operator whether to support a common core functional set, and certainly what extended functionality to provide as well. But a common subset for the extended community of operators will enable cross-pollination among them.

Confidentiality, Availability and Integrity of Experiment Ecosystem: The short-term focus for both existing facilities, and those in development during the near term, should be the implementation of basic protection mechanisms, and interfaces to the control them. The likely focus should be on experimenter data assets, providing users with the ability to limit visibility of sensitive assets, and controls on sharing assets generally to be within a set of users controlled by workgroup leader. Basic policy models should be supported for a common set of roles and privileges. Administrative interfaces should be provided for owner/operators to manage them.

Operational integrity of operational experiments should also be addressed with development of protection mechanisms for operational experiments, access controls that use these mechanisms, policy models for defining the access controls, programmatic interfaces to defining policies, and security administration user interfaces for operators and users.

Availability should be addressed at least at the level of policy modeling and meta-experimentation with a variety of resource management schemes.

Social and Cultural Changes: In the near term, the basis for social/cultural change will need to be created with a much greater degree of engagement with CEF users. Particularly in the area of usability, user-centered design and design testing will require periodic workshop-style interaction with a careful cross-section of users. The same is true in terms of design for workflow and common experimental methodologies.

Regarding existing practice, research in CEF infrastructure will need to give CEF operators the ability to begin to build a library collection of reusable experiment components, and use common cataloging capabilities. Co-development with short-term advances in frameworks and building blocks (see Section 4.3) will be required for the library items to be broadly usable. Early engagement with specific domains (see Section 4.1) will be important for a broad scope of social/cultural change.

Mid Term – 5 Years

Usability: In the mid term, meta-research in CEF human factors should continue on the basis of previous work, but in some cases with an additional focus on scientific usability test and other later-stage commercial best practices for software usability. In some areas – perhaps in asset management – short-term work will have matured sufficiently for this change in focus. In other areas – perhaps availability and resource management policy and administration – early work will have focus on modeling and policy management, with implementation starting later.

In addition, mid term work should transition into domain-specific usability support. With a basis in user interfaces for CEF usage and administration common across domains, resulting from short-term work, domain-specific extensions can support a particular research

domain's terminology, workflow, and methods typically preferred within an established community of interest.

Confidentiality, Availability and Integrity of Experiment Ecosystem: In the mid term, two additional areas should come into play. First, with basic confidentiality, integrity, and related access control functionality already established, the assurance of these mechanisms should become a focus. Second, availability should become a significant focus – including but not limited to mechanisms, usage, administration, and human interfaces for scheduling and for administration – especially for managing the usage of experimentation resources for operational experiments.

In terms of assurance, the initial focus should be on incorporating commonly used system components, e.g., SE Linux, into implementation of core infrastructure components.

Social and Cultural Changes: Mid term work must include activities specifically to build awareness of CEF capabilities, and the flexibility of them created in short-term work in other areas of experimental research. Flexibility and extensibility will be critical pre-requisites to engaging users to make contributions back to the research community: initially experiment components in libraries but also experimenter tools that may become new CEF capabilities.

Two other factors will be required to continue to spin up the scale of engagement with and collaboration among CEF users. Existing libraries and other sources of re-usable experiment components and tools – especially those warehoused outside of CEF infrastructure in open source repositories – must be the focus of expanding search capabilities to match experimenter needs with existing reusable artifacts. Also, some research funding vehicles should require not only the use of CEF infrastructure, but also deliverables that are useful for other researchers, especially in domain specific work.

Long Term – 10 Years

Usability: Long-term work should focus on flexibility and extensibility of designs and implementations of human interfaces of CEF workflows and models for experimentation and administration. We should expect some convergence of mid-term work in both usability and in frameworks and building blocks: common technical interfaces and software building blocks for common CEF capabilities and human interfaces to them.

With the resulting flexible architectures, long-term work can focus on techniques to easily adapt to domain specific needs, and variation for different levels of user sophistication. While domain-specific work will have been done in the mid term, we can also expect an expansion in the applicable scope of cyber-experimentation and research infrastructure to support it: more domains, new workflows, new ontologies, new methodologies, etc. Agile human interface technology will be essential to keep the pace.

Finally, with a sound basis of scientifically tested usability of workflow and methodology – and the user experience implementation of them – we should also expect developments in human interfaces designed for training user/experimenters, and accelerating the growth of researchers trained in CEF usage. Initial results in this area will likely be in support of common CEF features, processes, and procedures; but extension to domain-specific methodologies should follow.

Confidentiality, Availability and Integrity of Experiment Ecosystem: Near to mid term work should have applied today's well-understood concepts of confidentiality, integrity, availability, and related security functionality and concepts to CEF infrastructure. Experience in that time frame should have discovered new kinds of security requirements and features desirable for CEF infrastructure, and new applications of security concepts to the evolving ontologies and implementations of CEF infrastructure broadly. Research in the longer term should address design, implementation, and usage of CEF security measures to meet those to-be-emerging needs.

With availability in particular, we can expect considerable evolution as the scale of CEF infrastructure increases. Where medium availability features may have been sufficient – for example, tools and methods for scheduling and resource management – new models, tools, and capabilities will be needed in the longer term.

Social and Cultural Changes: Short to mid term activities will have to create substantial social and cultural changes from current norms. These changes are so substantial that concrete activities in the long term are difficult to predict. However, if short to mid term activities are successful in engendering the changes described above, then longer term changes should be characterized by a significant reversal of the localized testbed concepts of the recent past: researchers in many domains will find CEF capabilities so useful and familiar that they will specifically seek to use existing CEF infrastructure rather than invent their own environments for cyber research activities.

One critical factor in all time frames, continuing into the long term, is the broadening use of CEF infrastructure for advanced undergraduate and graduate education, and graduate research. As CEF infrastructure use becomes more routine in the academic environment, then more academic researchers will regard CEF usage as an expected part of cyber research, over the longer term.

5 CONCLUSIONS AND COMMUNITY RECOMMENDATIONS

This report presented a strategic plan and roadmap intended to catalyze generational advances in the field of experimental cybersecurity research. These results represent the conclusions of a study conducted under NSF auspices by SRI International and USC-ISI throughout calendar year 2014. The study had broad participation by stakeholders representing the cybersecurity research, research sponsor, and customer communities. The report outlines the process and methodology of the project, presents key inputs, supporting evidence developed through the course of the study, and synthesized results.

The overarching finding of this effort is that transformational progress in three distinct, yet synergistic, areas is required to achieve the desired objectives:

1) Fundamental and broad intellectual advances in the field of experimental methodologies and techniques, with particular focus on complex systems and human-computer interactions.
2) New approaches to rapid and effective sharing of data and knowledge and information synthesis that accelerate multi-discipline and cross-organizational knowledge generation and community building.
3) Advanced, accessible experimentation infrastructure capabilities.

While each of these is not surprising, the important recommendation is that significant investment, R&D and coordination is required in all three areas simultaneously. The most successful testbed projects have been those that were pursuing a rigorous research agenda both in terms of the meta-research into testbeds themselves and in close collaboration with the users of the facilities.

The central result of our study is a roadmap that presents requirements, objectives and goals in each of the areas outlined above over three, five and ten year phases. In some case the phases build upon each other, and in other cases, new fundamental research is required over a long period of time to satisfy the objectives of the roadmap.

The roadmap described a set of 30 key capabilities in eight core areas envisioned for realizing a broad agenda of research, development and infrastructure. These capabilities are key to realizing a national strategy for cybersecurity experimentation of the future. Each of the capabilities of the roadmap describe an important class of functionality, however it is important to look holistically across all of the capabilities to paint a comprehensive picture of fundamental research, community and advanced infrastructure.

5.1 ROADMAP FINDINGS

Section 4.1 of the roadmap discussed *Domains of Applicability* and described the multi-year roadmap elements in the areas of:

- Support for cross domain experimentation (critical infrastructure sectors)
- Multidisciplinary experimentation including computer science, engineering, mathematics, modeling, human behavior, sociology, economics, and education

- Portability of experiments, packaged for sharing and use in cross-discipline experiment

Section 4.2 of the roadmap discussed *Modeling the Real World for Scientifically Sound Experiments* and described the multi-year roadmap elements in the areas of:

- Models of real world environments
- Experiments that scale
- Experimentation with systems-of-systems
- Human activity

Section 4.3 of the roadmap discussed *Frameworks and Building Blocks for Extensibility* and described the multi-year roadmap elements in the areas of:

- Workflow and management of experiment lifecycle
- Open interface APIs
- Libraries and other building blocks
- Integration frameworks, tools, and designs

Section 4.4 of the roadmap discussed *Experiment Design and Instantiation* and described the multi-year roadmap elements in the areas of:

- Design tools, specifications, ontologies, and compilers
- Reusable designs for science-based hypothesis testing
- Automated discovery of local and distributed resources
- Dynamic instantiation of domain-specific test apparatus
- Validation of instantiated test environments and apparatus

Section 4.5 of the roadmap discussed *Interconnected Research Infrastructure* and described the multi-year roadmap elements in the areas of:

- Automation of interconnection that enables sharing resources
- Interconnection that can be dynamic or on-demand
- Support for a variety of types of experiments

Section 4.6 of the roadmap discussed *Experiment Execution and Management* and described the multi-year roadmap elements in the areas of:

- Experiment orchestration
- Visualization and interaction with experiment process
- Experiment debugging with checkpoint and rollback
- Experiment execution validation

Section 4.7 of the roadmap discussed *Instrumentation and Experiment Analysis* and described the multi-year roadmap elements in the areas of:

- Instrumentation and data collectors
- Transport and protection mechanisms
- Data repositories
- Data analysis

Section 4.8 of the roadmap discussed *Meta-Properties* and described the multi-year roadmap elements in the areas of:

- Usability (experiments, owner/operator)
- Confidentiality, availability and integrity of experiment ecosystem
- Social and cultural changes

All of the roadmap sections highlight initial steps that can be taken in the next three years to jumpstart the community and create an initial, usable, shared capability. Also highlighted are advanced R&D and fundamental long-term research to be undertaken over the next five to 10 years. It is clear that there are many existing capabilities that can be leveraged and that yet even more coordination, collaboration and sharing is required in order to achieve the necessary advances to realize the capabilities identified in the roadmap.

5.2 CONCLUSION

The capabilities identified in the roadmap take into account the current state of the art in experimental cybersecurity research and its supporting infrastructure. We found that a large amount of the research infrastructure and capabilities needed either do not exist or are not generally available for use. A set of shared, vetted community research capabilities will provide a solid basis upon which to build future cyber experimentation environments. Its use will increase sharing amongst researchers and reduce the time and money spent building one-off test environments in support of new research efforts. Building upon previously vetted capability components and utilizing test environment validation tools will improve overall experimental result quality. Uncaught errors in a test environment can not only invalidate a specific experiment but could also amplify as other research efforts are built upon invalid results and then make their way into products.

The results of this study strongly point to the following three conclusions:

1) A new generation of experimental cybersecurity research is needed that will help shift the asymmetric cyberspace context to one of greater planning, preparedness, and higher assurance fielded solutions.

2) Emphasis on equipment and related software infrastructure alone will fall far short of achieving the transformational shift in research, community, and supporting experimentation required to address cybersecurity in the rapidly changing cyber environment. In addition to leveraging current and expected capabilities in cybersecurity and adjacent areas, we assume there will be advances in key computer science disciplines such as ontologies, meta-data, libraries, and corresponding resource discovery.

3) Strong, coupled, and synergistic advances across three areas – fundamental methodological development, fostering and leveraging communities of researchers, and advancing capabilities of the infrastructure supporting that research – will move the field beyond today's state of the art.

6 ACKNOWLEDGEMENTS

The CEF team thanks the National Science Foundation (NSF) and especially Kevin Thompson and Anita Nikolich of the Advanced Cyberinfrastructure (ACI) Division for their support.

The team greatly appreciates the guidance received from the members of the Advisory Group (see Appendix C), as well as their comments and feedback on the draft report.

We thank all the members of the community who reviewed and provided comments and feedback on the draft report.

The team thanks John Wroclawski and Stephen Schwab at USC-ISI for helping review and synthesize study group outputs and to frame and write the report. We also thank John Sebes, a consultant to USC-ISI, for helping write the roadmap. We thank Matt Binkley and Joe Kemp for helping run Study Group 2 and other meetings.

The CEF team thanks Linda Briesemeister, Matt Binkley, Ted Faber, and Goran Scuric for help with the assessments of existing experimentation infrastructure. We thank Mary Denz for hosting a visit to the Air Force Research Laboratory and Lee Rossey for hosting a visit to MIT Lincoln Laboratory.

The team especially thanks all of the study group participants (see Appendix B.4) for taking the time and effort to participant in the multi-day meetings and share their knowledge and insight into future cybersecurity challenges and experimentation needs. We also thank Brian DeCleene, Stephen Schwab, David Corman, Alefiya Hussain, George Kesidis, Zachary Tudor, Ritu Chadha, Sami Saydjari, and Dale Johnson, for helping to run the study groups by facilitating breakout group sessions.

7 REFERENCES

[1] Protected Repository for the Defense of Infrastructure Against Cyber Threats (PREDICT), U.S. Department of Homeland Security Science and Technology Directorate, http://www.predict.org/.

[2] The DETER Project, USC Information Sciences Institute, http://www.deter-project.org/.

[3] Global Environment for Network Innovations (GENI), Raytheon BBN Technologies, http://www.geni.net/.

[4] The National Cyber Range (NCR), Defense Advanced Research Projects Agency, https://www.whitehouse.gov/files/documents/cyber/DARPA%20-%20NationalCyberRange_FactSheet.pdf.

[5] National Cyber Range Overview, Test Resource Management Center, Office of the Secretary of Defense, February 24, 2015, http://www.acq.osd.mil/dte-trmc/docs/20150224_NCR%20Overview_DistA.pdf.

[6] Cyber Security and Information Assurance Interagency Working Group (CSIA IWG), Networking and Information Technology Research and Development (NITRD) Program, https://www.nitrd.gov/nitrdgroups/index.php?title=Cyber_Security_and_Information_Assurance_Interagency_Working_Group_%28CSIA_IWG%29.

[7] IFIP Working Group 10.4 on Dependable Computing and Fault Tolerance, http://www.dependability.org/wg10.4/.

[8] Institute for Information Infrastructure Protection (I3P), http://www.thei3p.org/.

[9] DeterLab: Cyber-Security Experimentation and Testing Facility, USC Information Sciences Institute, http://www.isi.deterlab.net

[10] ProtoGENI, U.S. National Science Foundation, http://www.protogeni.net/.

[11] Open Networking Lab (ON.lab), http://onlab.us/.

[12] Connected Vehicle Test Beds, U.S. Department of Transportation, http://www.its.dot.gov/testbed.htm.

[13] Mcity Test Facility, Mobility Transformation Center, University of Michigan, http://www.mtc.umich.edu/test-facility.

[14] Cyber Experimentation Environment (CEE), Air Force Research Laboratory (AFRL).

[15] Cyber System Assessment Projects, MIT Lincoln Laboratory, https://www.ll.mit.edu/mission/cybersec/CSA/CSA-projects.html.

[16] Cyber-Security Collaborative Research Alliance, Pennsylvania State University, http://cra.psu.edu/.

[17] Cyber Security Research Alliance, U.S. Army Research Laboratory, http://www.arl.army.mil/www/default.cfm?page=1417.

[18] Science of Security (SoS) Lablet, Information Trust Institute, University of Illinois, http://www.iti.illinois.edu/research/evaluation/science-security-sos-lablet.

[19] Coordinating User and Technical Security (CUTS), Human and Technical Security (HATS), University of Indiana, http://usablesecurity.net/CUTS/cuts.php.

[20] Trustworthy Cyber Infrastructure for the Power Grid (TCIPG), Information Trust Institute, University of Illinois, http://tcipg.org/.

[21] Foundations Of Resilient CybEr-physical Systems (FORCES), University of California, Berkeley, https://www.cps-forces.org/.

[22] Correct-by-Design Control Software Synthesis for Highly Dynamic Systems, University of Michigan, http://www.dynamiccps.org/.

[23] The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research, Technical Report, U.S. Department of Homeland Security, August 2012, https://www.caida.org/publications/papers/2012/menlo_report_actual_formatted/.

[24] Applying Ethical Principles to Information and Communication Technology Research: A Companion to the Menlo Report, Technical Report, U.S. Department of Homeland Security, October 2013, http://www.caida.org/publications/papers/2013/menlo_report_companion_actual_formatted/.

[25] Opnet, Riverbed Technology, http://www.riverbed.com/products/performance-management-control/opnet.html.

[26] Mechanical Turk, Amazon, https://www.mturk.com/mturk/welcome.

[27] Lincoln Adaptable Real-time Information Assurance Testbed (LARIAT), MIT Lincoln Laboratory, https://www.ll.mit.edu/mission/cybersec/CSA/CSA-projects.html#lariat.

[28] KVM-based Zero (0) Artifact LARIAT Agent (K0ALA), https://www.ll.mit.edu/mission/cybersec/CSA/CSA-projects.html#k0ala.

[29] Skaion Corporation, http://www.skaion.com/.

[30] DETER Agents Simulation Humans (DASH), USC Information Sciences Institute, http://deter-project.org/deterlab_software.

[31] The Information Design Assurance Red Team (IDART™), Sandia National Laboratories, http://www.idart.sandia.gov/.

[32] PlanetLab, Princeton University, https://www.planet-lab.org/.

[33] Emulab, Flux Research Group, School of Computing University of Utah, https://www.emulab.net/.

[34] Network simulator 2 (ns-2), USC Information Sciences Institute, http://www.isi.edu/nsnam/ns/.

[35] Montage AGent Infrastructure (MAGI), USC Information Sciences Institute, http://montage.deterlab.net/magi/index.html.

[36] SETI@Home, University of California, http://setiathome.ssl.berkeley.edu/.

[37] VMware, Inc., http://www.vmware.com/.

[38] Xen Project, Linux Foundation, http://www.xenproject.org/.

[39] pcap (packet capture), Wikipedia, https://en.wikipedia.org/wiki/Pcap.

[40] syslog (system logging), https://en.wikipedia.org/wiki/Syslog.

[41] Windows Event Log, Microsoft, https://msdn.microsoft.com/en-us/library/windows/desktop/aa385780%28v=vs.85%29.aspx.

[42] System for Internet-Level Knowledge (SiLK), CERT Network Situation Awareness Team, Software Engineering Institute, Carnegie Mellon University, https://tools.netsa.cert.org/silk/.

[43] Audit Record Generation and Utilization System (Argus), QoSient, LLC, http://qosient.com/argus/.

[44] The Bro Network Security Monitor, The Bro Project, https://www.bro.org/.

[45] Snort, Cisco Systems, https://www.snort.org/.

[46] Route Views Project, Advanced Network Technology Center, University of Oregon, http://www.routeviews.org/.

[47] Routing Information Service (RIS), Network Coordination Center, Réseaux IP Européens (RIPE), https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris.

[48] Tcpdump & Libpcap, http://www.tcpdump.org/.

[49] Wireshark, Wireshark Foundation, https://www.wireshark.org/.

A SURVEY OF EXISTING EXPERIMENTATION INFRASTRUCTURE

The initial survey set consisted of 46 candidates. Candidates were:

1. Aerospace Corporation Cybersecurity Testbeds
2. Air Force Research Laboratory Cyber Experimentation Environment (CEE)
3. Applied Research Associates Cyber Test Bed (CTB)
4. ART FLIGHTLAB
5. Battelle – Cloud Computing
6. Battelle – Networking
7. Battelle – QKD Test Bed
8. UC Berkeley Building-to-Grid (B3G) Testbed
9. Carnegie Mellon University (unspecified)
10. National Cyber Range (NCR)
11. USC-ISI DETERLab
12. DoD Discrete Event Simulation
13. DoD Tactical Network Testbed (TNT)
14. Tactical Network Testbed (TNT) Experiments
15. DoE Cyber Physical Systems / Power
16. U.S. Department of Transportation (DOT) Connected Vehicle Test Beds
17. European Union FIRE
18. National Science Foundation (NSF) GENI
19. George Mason University (GMU) OCTANE
20. Honeywell Programmable Router Testbed
21. Iowa State PowerCyber Testbed
22. Idaho National Laboratory Wireless Range
23. Idaho National Laboratory / Sandia National Laboratory National SCADA Testbed
24. University of Illinois Trustworthy Cyber Infrastructure for the Power Grid (TCIPG)
25. Intel Open Cirrus
26. Iowa State's ISEAGE
27. Merit Network's Michigan Cyber Range
28. University of Michigan Transportation Research Institute (UMTRI) Range
29. MIT Lincoln Laboratory LARIAT
30. Japanese National Institute of Information and Communications Technology (NICT) JGN2plus
31. NIST Smart Grid Testbed
32. NIST Factory Equipment Network Testbed (FENT)
33. Northrup Grumman's Cyber Test Range
34. NSA / CSS Global Information Grid (GIG) Testbed
35. NSF Workshop on Network Research Testbeds
36. Ohio University The Boat of Knowledge in Science (BooKS)
37. Pacific Northwest National Lab (PNNL) PowerNET
38. Princeton's PlanetLab
39. Rutgers University Orbit Wireless Network Testbed
40. Sandia National Laboratory Advanced Systems Test Beds
41. Sandia National Laboratory Firewheel
42. Skaion Corporation Traffic Generator
43. Southwest Research Institute (SwRI) Connected Vehicle Affiliated Test Bed
44. Stanford and Berkeley Open Networking Lab (ON.Lab)
45. University of Buffalo PhoneLab
46. University of Illinois Wireless Wind Tunnel (iWWT)

After an initial round of Internet-based research, this set was reduced by merging duplicate efforts and by removing test infrastructures for over-represented domains, those that were not applicable, and those for which no information could be found.

The final survey set was then divided amongst a team of researchers with testbed experience at both USC-ISI and SRI. Each researcher interviewed one or more key people at the owning organization and filled out a survey template. In some cases, in-person visits were made to see the infrastructure in operation in its native environment. A summary table plus the completed surveys follow.

Table 18. Representative cybersecurity experimentation infrastructures surveyed under CEF considered multiple sources and domains.

Section	Test Infrastructure	Domain	Availability
A.1	Air Force Research Laboratory Cyber Experimentation Environment (CEE)	Information Systems and IT Networking	US Gov't only
A.2	USC-ISI DeterLab	Information Systems and IT Networking	Open
A.3	Department of Transportation Connected Vehicle Test beds	Transportation (Automotive)	Open
A.4	European Union FIRE Initiative	IT Networking	Open
A.5	National Science Foundation GENI	Information Systems and IT Networking	Open
A.6	NICT Japan StarBed[3] / JGN2+	IT Networking	Open
A.7	MIT Lincoln Laboratory (MIT-LL)	Information Systems and IT Networking; DoD	Transition to open in process
A.8	Test Resource Management Center (TRMC) National Cyber Range (NCR)	Information Systems and IT Networking	U.S. Gov't only
A.9	George Mason University (GMU) OCTANE	Transportation (Automotive)	Open
A.10	Open Networking Lab (ON.Lab) Software Defined Networking (SDN) Testbed	IT Networking	Contributors only
A.11	Intel Labs Open Cirrus	Cloud Computing	Affiliates only
A.12	Rutgers ORBIT	IT Networking (Wireless)	Open
A.13	University of Buffalo PhoneLab	Communications (Android)	University affiliates only
A.14	Iowa State PowerCyber	Energy (Electric Power)	Open
A.15	Pacific Northwest National Laboratory (PNNL) PowerNet	Energy (Electric Power)	Requires approval
A.16	Skaion Traffic Generator	Information Systems and IT Networking	Open, with fee
A.17	Department of Energy (DOE) TCIPG	Energy (Electric Power)	Affiliates only
A.18	Applied Communications Sciences (ACS) Virtual Ad Hoc Network (VAN)	IT Networking (Wired & Wireless)	Open, with fee

A.1 AIR FORCE RESEARCH LABORATORY CYBER EXPERIMENTATION ENVIRONMENT (CEE)

INTRODUCTION

Primary domain(s)	Information Systems and IT Networking
High Level Description	The AFRL Cyber Experimentation Environment (CEE) is a reconfigurable network testbed that supports Department of Defense funded cybersecurity research within a classified environment. The testbed has many advanced capabilities, comprised of both tools acquired from outside vendors and tools developed inside the classified network. Those acquired from vendors may be accessed by contacting the vendors directly, however those developed inside the classified test environment may not be extracted for use.
Maturity level	Relatively mature.

LOGISTICS

Website URL	N/A
Owning Organization(s)	Air Force Research Laboratory (AFRL), Rome, NY
Funding Source(s)	United States Air Force

A.2 USC-ISI DETERLAB

INTRODUCTION

Primary domain(s)	Information Systems and IT Networking
High Level Description	The DETER testbed is a reconfigurable network testbed that supports many researchers, each of whom may simultaneously run one of several experiments within a custom-specified experiment topology of their own design. The DETER testbed space-shares the available physical resources. The testbed mainly supports researchers and higher education coursework laboratory experimentation. The typical experiment is performed using PCs or virtual machines configured to play the role of clients, routers, middle-boxes, or servers, interconnected by networks with researcher-specified characteristics in experimental scenarios.
Maturity level	Age: 10 years Additional Info: Advanced experimentation facility with a primarily advanced-user researcher/experimenter demographic. New features are incrementally introduced.

LOGISTICS

Website URL	http://deter-project.org/ (public-facing site) https://www.isi.deterlab.net/index.php3(testbed account and resources) https://education.deterlab.net (education site) https://trac.deterlab.net (wiki site) http://www.cyber.st.dhs.gov/deter/
Owning Organization(s)	Department of Homeland Security (DHS) Science and Technology Directorate University of Southern California (USC) Information Sciences Institute (ISI) (Marina del Rey, CA) University of California (UC) Berkeley (Berkeley, CA)
Funding Source(s)	Department of Homeland Security (DHS), National Science Foundation (NSF), Defense Advanced Research Projects Agency (DARPA)

USAGE

Research problems best used for	DETER is a research project for which DeterLab is the facility proving ground that provides the needed infrastructure and advanced tools to perform cybersecurity experimentation, within an emulated internet environment, for *researchers to develop, test, and discover scientifically-based cybersecurity innovations that lead to significant increases in scale, pace, and power.* DeterLab is a shared testbed providing a platform for research in cybersecurity and serving a broad user community, including

	academia, industry, and government. To date, DeterLab-based projects have included behavior analysis and defensive technologies including DDoS attacks, worm and botnet attacks, encryption, pattern detection, and intrusion-tolerant storage protocols.
Usage Restrictions	Open and Free but applications must be vetted by the DETER Project Staff to ensure best practices concerning security of assets.
Access Method	Internet web based access
Approximate number of user organizations	100's of organizations, spanning 30 countries, 10 governments, 206 universities, 31 institutes, 19 laboratories, 51 companies, and 8 classified as "other".
Approximate number of users	Currently, there are 3,788 users at 468 locations (where location is equivalent to a city.) DeterLab has grown into a facility where over 600 researchers have conducted network and cyber-security experimentation. DeterLab users have conducted hundreds of research projects and published more than 100 resulting scientific papers. More than 3800 students, the crucial next cyber-security generation, have received hands-on training via DeterLab.

INFRASTRUCTURE NATURE

The DETER testbed is a facility for repeatable, controlled experiments. It provides for:

- Experiment isolation and perimeter security
- Clean room repeatable experimentation
- Controlled outside interaction/observation
- Resources allocated on demand
- Scheduled experiments can be supported
- Education resource
- Built on top of Emulab Software

DETER provides research and experimentation support through:

- Use automation to manage experiments through their complete lifecycle,
- Model and construct experiments,
- Run and monitor experiments,
- Adjust their scale and resolution, and
- Gather and analyze experimental result data.

DETER testbed subsystems include:

- Assignable resources including physical and virtual nodes an links
- *Containers*, for multi-resolution virtualization of experiment resources. For creating large-scale DETER topologies that support differing degrees of fidelity in individual elements.
- *DASH*, for predictive modeling of human behavior supporting definition of mental models, reactive goal-driven behavior, and combinations of deliberative/instinctive behaviors.

- *Federation* and its *ABAC* authorization library, for connection of heterogeneous resources from different owners with varying usage and access policies.
- *ABAC* - Scalable authorization system used with the federation capability.
- *Multi-party experiments technology* that provides controlled but co-joined experiments, creating different views of the experiment environment for multiple experimenters or groups of experimenters within one unified experiment.
- *MAGI*, for orchestrating networking experiments providing deterministic control over the various components in an experiment.
- Various Tools for experiment control, management and behavior control

Experiments can scale to thousands of nodes, including both physical and virtual nodes.

INFRASTRUCTURE OPERATION

DETER has successfully supported projects and experiments related to:

- Utility oil/gas/power system security testing
- DDOS attack testing
- Worms and Botnet experimentation
- Education courses related to various aspects of cybersecurity
- Smart Grid security experimentation
- Power outage simulations
- Etc.

DETER also provides for:

- Risky Experiment Services
- Controlled permeability of the DETER isolation

EXPERIMENT LIFECYCLE SUPPORT

Deter provides various tool and APIs for complete experiment lifecycle support. It includes rich set of web based APIs and tools to:

- Create and manage user accounts, projects an experiments
- Model and construct experiments,
- Graphical Experiment Editor
- Run, control and monitor experiments,
- Adjust their scale and resolution, and
- Gather and analyze experimental result data.
- Sharing of experiment data
- Low Level Experiment Access
  - Node Access
  - Node Control
  - Network information
- Problem reporting and tracking system
- MAGI system for orchestrating networking experiments in DETERLab.
  - Enables deterministic orchestration of **event streams**
  - Repeatable enactment of **procedure**
  - Provides **control** and **data management** for experiments

EXAMPLE EXPERIMENTS THAT USED THIS INFRASTRUCTURE

First 20 in alphabetic order:

- A Classification Layer for new network Architectures, University of California, Berkeley
- A dos-limiting network architecture, UC Irvine
- A large-scale measurement for malware analysis, University of California, Berkeley
- A network security laboratory to enable senior level students to apply their theoretical knowledge using the offered experiments, Jordan University of Science and Technology
- A platform for distributing computation onto volunteer nodes, USC
- A secure DDoS protection framework, The University of Tokyo
- Addressing TCP limitations in public clouds, Berkeley Computer Science
- Advanced Computer Security, National University of Science and Technology
- Advanced Persistent Threat Classification, Barnstormer Softworks, Ltd.
- An experimental distributed deterministic OS, Yale University
- Analysing traffic patterns in modified BitTorrent P2P clients Universidad de Buenos Aires
- Anonymous messaging protocols and systems, Bar Ilan University, Israel
- Application Layer DDoS Attack Detection Attribution and Mitigation System, UCSB
- Applied Cryptography and Network Security Course, Radford University
- Architecture of Secure Operating Systems, CUNY John Jay College
- Attack classification, Thiagarajar College of Engineering
- Cybersecurity of oil/gas pipeline SCADA, Southwest Jiaotong University

A.3 DEPARTMENT OF TRANSPORTATION CONNECTED VEHICLE TEST BEDS

INTRODUCTION

Primary domain(s)	Transportation (Automotive)
High Level Description	Connected Vehicle Test Beds are real-world, operational test beds that offer the supporting vehicles, infrastructure, and equipment to serve the needs of public- and private-sector testing and certification activities. The main purpose of these testbeds is to explore the overall architecture of the wheeled vehicle transportation system.
Maturity level	Mature and growing. Includes vehicle to vehicle and vehicle to infrastructure communication system.

LOGISTICS

Website URL	http://www.its.dot.gov/testbed.htm
Owning Organization(s)	U.S. Department of Transportation (US DOT)
Funding Source(s)	U.S. Department of Transportation

USAGE

Research problems best used for	Testing vehicle safety, mobility, and environmental applications, services, and components, including wireless communication between vehicles. http://www.its.dot.gov/factsheets/pdf/ConnectedVehicle_testbed.pdf
Usage Restrictions	None, open to anyone. US DOT will provide limited assistance and equipment may be loaned out. Data provided free of charge. Users must pay costs for testing that requires human drivers.
Access Method	In person after approval.

INFRASTRUCTURE NATURE

This is not a laboratory. These are real vehicles with special equipment, driven on real roads that have real electronic/communications based traffic signs. Data from the vehicles is sent to back end servers and made available to researchers and users.
The scale supported is based on the number of units that can be equipped, where “unit” means any “thing” in the system. Believes it can be scaled to potentially 10’s of thousands of vehicles.

INFRASTRUCTURE OPERATION

The test infrastructure is a mini version of the real transportation system they anticipate deploying, which operates in the real world. They break the transportation system into four classes of objects: moving objects (e.g., vehicles), carried objects (e.g., Bluetooth devices, CDs, etc.), fixed boundary objects (e.g., traffic signs, etc.) and Internet connected back end systems. They experiment with the system to determine what the safety requirements should be.

There is special instrumentation that can be added to objects to collect additional data. The data is made available for use by researchers through the US DOT research data exchange.

EXPERIMENT LIFECYCLE SUPPORT

They do not have any experiment lifecycle support, other than data collection on the backend side. Their mind set is focused more on designing the overall transportation system and giving out specs to vendors for developing components that will plug into the system. They would like a CAD design tool that can be used to reflect a transportation system design, but they don't have anything like that now.

EXAMPLE EXPERIMENTS THAT USED THIS INFRASTRUCTURE

Tests and experiments to analyze the overall transportation system design and identify requirements for fortifying device boundaries to make them less susceptible to attack.

Data analysis for integrity attacks: Vanderbilt is looking at information flows in the data to determine if the flows be altered.

Experiments to determine the efficacy of approaches to solve privacy issues with mandated tracking devices (via the National Highway Traffic Safety Administration (NHTSA)). One approach involved changes in devices so that they provide no identifying information such as fixed mac addresses or IP addresses.

In 2015, they are planning a large-scale deployment trial (3000+ vehicles) in Ann arbor to examine crash avoidance technology. All data is collected and stored. One day's worth of data from a prior deployment trial was made available via their research data exchange. They invited individuals to try to identify the owner and individual car using the data.

A.4 EUROPEAN UNION FIRE INITIATIVE

INTRODUCTION

Primary domain(s)	IT Networking
High Level Description	European Union Future Internet Research & Experimentation (FIRE) Initiative A variety of network experimentation infrastructures and tools with different technologies and characteristics. Various structures, tools and features are already available and trials are being performed. All of the facilities evolve in a demand-driven way, supported through open calls. BonFIRE (Clouds), OFELIA (OpenFlow) and CREW (Cognitive Radio) are three examples of FIRE facilities now offering Open Access; other individual testbeds federate with running FIRE Facility projects. The testbed hardware is owned by individual institutes (hardware is typically not funded by the European Commission) and through projects the APIs and federations are developed. Fed4FIRE (www.fed4fire.eu) is a federation project combining multiple FIRE facilities with the same set of APIs, which are compatible with the GENI APIs. The frameworks of BonFIRE, OFELIA, CREW, PlanetLab, etc. are federated through Fed4FIRE. Fed4FIRE runs currently both with funded access for experimenters as free open access.
Maturity level	Mature. They have published their final roadmap report.

LOGISTICS

Website URL	http://www.ict-fire.eu/home.html
Owning Organization(s)	European Union (EU)
Funding Source(s)	European Commission under FP7 ICT Objective 1.6 Funding period has almost ended. Local government and institutes typically fund the hardware, while software stacks and open calls for experimenters are funded by the EC. Fed4FIRE is one of the last funded project of this Objective 1.6 and is funded till end of 2016. The HORIZON 2020 framework is the successor funding framework of the EU.

USAGE

Research problems best used for	For experimentation with networks that require a multi-disciplinary test environment for validating highly innovative and revolutionary ideas for new network and service paradigms.
Usage Restrictions	Open and public for all experimenters who find the facilities offered are suited to their R&D needs. Use and support are free.
Access Method	Remote supported

INFRASTRUCTURE NATURE

FIRE is a federation of network testbeds and a unified interface to interconnect them. The federation supports a group of associated research projects carrying out network experiments.

While some testbeds offer their infrastructure code for download, the primary model is that researchers remotely access resources from one or more federated testbeds.

A representative sample of testbeds includes:

The BonFIRE and OFELIA testbeds are completely open for access. BonFIRE provides multi-site cloud-based infrastructure as a service. OFELIA is an OpenFlow based networking testbed.

Fed4FIRE specifies and provides distributed access to multiple EU testbeds to enable via common interfaces and interconnectivity (also with the US, Brazil, Japan, Korea). As individual institutes own the testbed hardware, the institutes determine the policies. However, most facilities have currently an open access policy. Testbeds as BonFIRE, OFELIA, CONFINE, CREW, PlanetLab, etc. are currently federated through Fed4FIRE which offers funded Open Calls for experiments, funds for co-operation with the U.S. and GENI (with matched SAVI funding in the U.S.) and also Open Access.

CONFINE provides access to community-owned IP networks. Access is more limited.

CREW is a spectrum-sensing, cognitive radio testbed. It allows Open Access at a best effort level. Dedicated experimenters can be vetted for more support.

EXPERIMEDIA supports development of multimedia experiments. The facility is moving to an open access framework.

OpenLab interconnects multiple networking testbeds, including PlanetLab Europe. Researchers can petition for access.

FLEX and SUNRISE are just coming on line. FLEX is a cellular access testbed and SUNRISE supports ocean sensing.

The scale of experimentation varies some by testbed, but generally testbeds support a few hundred nodes. Some – especially those exporting virtual machines – support a few thousand.

INFRASTRUCTURE OPERATION

Generally the networking resources are physical resources and the computation resources are cloud-based virtual machines or physical resources deployed in Emulab style testbeds. Again, there is some variation between the individual testbeds.

EXPERIMENT LIFECYCLE SUPPORT

Again, there is some variation between testbeds, but generally the systems automate allocation and configuration of resources. Experimenters operate and manage their own experiments. In Fed4FIRE, the lifecycle support and APIs, are compatible with the GENI framework. There is a good cooperation between Fed4FIRE and GENI (co-organized workshops, specification of common standards, federation on resources and dataplane connectivity, co-organized summer schools)

EXAMPLE EXPERIMENTS THAT USED THIS INFRASTRUCTURE

Each of these is a research project associated with FIRE. More are available in the FIRE brochure at http://www.ict-fire.eu/home/publications.html

CLOMMUNITY – Uses CONFINE and cloud infrastructure to iteratively design and prototype community networks. Target communities can access and use the evolving networks.
EULER – Researchers design evolvative routing protocols and compare their performance against traditional routing (e.g., BGP) in controlled networks.
3D-LIVE – Using EXPERIMEDIA 3D-LIVE deploys multimedia and emerging media applications. Users access telemersive applications under controlled conditions.
ALIEN – Builds a hardware abstraction layer on top of OPHELIA's OpenFlow infrastructure. Different abstractions and levels of transparency can be compared in similar networks.
In the EU, there is typically a drive towards industry experiments, besides academic experiments. E.g., in Fed4FIRE, specific open calls and funding are launched towards SMEs (small companies).

A.5 NATIONAL SCIENCE FOUNDATION GENI

INTRODUCTION

Primary domain(s)	Information Systems and IT Networking
High Level Description	GENI, a virtual laboratory for exploring future internets at scale, creates major opportunities to understand, innovate and transform global networks and their interactions with society. Dynamic and adaptive, GENI opens up new areas of research at the frontiers of network science and engineering and increases the opportunity for significant socio-economic impact.
Maturity level	GENI will: – Support at-scale experimentation on shared, heterogeneous, highly instrumented infrastructure; – Enable deep programmability throughout the network, promoting innovations in network science, security, technologies, services and applications; and – Provide collaborative and exploratory environments for academia, industry and the public to catalyze groundbreaking discoveries and innovation.

LOGISTICS

Website URL	http://www.geni.net/ http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=501055 http://www.geni.net/wp-content/uploads/2011/06/GENI-at-a-Glance-1Jun2011.pdf
Owning Organization(s)	The GENI Project Office, managed by BBN Technologies, under the leadership of Mark Berman, Project Director.
Funding Source(s)	National Science Foundation (NSF)

USAGE

Research problems best used for	Understand global networks and their evolving interactions with society. Innovate at the frontiers of network science and engineering. Explore future internet architectures Research in Software Defined Networking Cloud Networking Large Scale evaluation of protocols Transform the science of network research and society at large. **Infrastructure Technology** The core concepts for the suite of GENI infrastructure feature: • **Programmability** – Researchers may download software into GENI-compatible nodes to control how those nodes behave.

	• **Virtualization and Other Forms of Resource Sharing** – Whenever feasible, nodes implement virtual machines, which allow multiple researchers to simultaneously share the infrastructure and each experiment runs within its own, isolated slice created end-to-end across the experiment's GENI resources. • **Federation** – Different parts of the GENI suite are owned and/or operated by different organizations, and the NSF portion of the GENI suite forms only a part of the overall 'ecosystem'. • **Slice-based Experimentation** – GENI experiments will be an interconnected set of reserved resources on platforms in diverse locations. Researchers will remotely discover, reserve, configure, program, debug, operate, manage, and teardown distributed systems established across parts of the GENI suite.
Usage Restrictions	None
Access Method	Internet
Approximate number of user organizations	Industry Participants: Corporations including AT&T, Arista, Battelle, CA, Ciena, Cisco, CNRI, Cobham, Big Switch, Fujitsu, Hewlett-Packard, IBM, Infinera, Juniper, Microsoft Research, NEC, Netronome, Nicira and Qwest are working with GENI academic teams across the United States to help build, integrate, and operate early prototypes of GENI.
Approximate number of users	Researchers: As of January 2011, awards were made to 83 academic/industrial teams for various projects to build, integrate, and operate early prototypes of the GENI virtual laboratory. As of October 2014 GENI has over 2700 users.

INFRASTRUCTURE NATURE

GENI provides compute resources that can be connected in experimenter specified Layer 2 topologies. Experimenters can set up custom topologies, protocols and switching of flows
Usage of WIMAX technology. Wireless nodes and Android handsets available. Collaboration to implement national-scale infrastructure: – Sliced and deeply-programmable – Incorporating OpenFlow/SDN switches, GENI Racks, etc. – High-speed (10-100 Gbps) – Distributed cloud (racks) for content - caching, acceleration, etc. GENI is meant to enable: – At-scale experiments – Internet-incompatible experiments

- Both repeatable and "in the wild" experiments
- 'Opt in' for real users
- Instrumentation and measurement tools

INFRASTRUCTURE OPERATION

GENI grows by GENI-enabling heterogeneous infrastructure through usage of federated aggregates providing various resources. Resources are advertised by aggregates via XML messages.

EXPERIMENT LIFECYCLE SUPPORT

A unified approach to the experimentation lifecycle – experiment specification, resource inquiries reservation, experiment execution and termination.

XML based resource specification allows extensions in resource types and other parameters.

GENI experiment control tools are used to create slices, add or remove resources to slices, and delete slices.

Tools supporting orchestration – however they are aggregate specific.

EXAMPLE EXPERIMENTS THAT USED THIS INFRASTRUCTURE

Various Company Engagements

Lots of tools released to aid others in experimentation

TECHNOLOGY AND USAGE DETAILS

User interacts with GENI using standard API called the *GENI Aggregate Manager API* or *GENI AM API*. API allows experimenters to:

- List the resources available at an aggregate,
- Request specific resources from the aggregate be allocated to their slices,
- Find the status of resources from the aggregate that are allocated to their slices, and
- Delete resources from their slices.

GENI spans multiple participant networks called aggregates. Aggregate providers are owners/operators of campus networks or research testbeds federated with GENI.

An aggregate is a software server that provides resources to clients based on the GENI aggregate manager API. Through that API an aggregate may advertise resources of different types (computation, storage, network) and accept requests for allocating these resources for a period of time.

The aggregate may provide virtual 'sliced' or non-virtual 'whole' resources to customers. An aggregate generates custom private internal network topologies per request, and participates in a process for generating cross-aggregate custom private network topologies known as stitching. Each aggregate advertises their resources. GINI resources are specified in RSpec documents in XML format.

Experimenters send to aggregates a *request RSpec* that describes the resources they want and get back from the aggregates a *manifest RSpec* that describes the resources they got. The manifest includes information the experimenters will need to use these resources such as the names and IP addresses of compute resources (e.g. virtual machines), user accounts created on the resources and VLAN tags assigned to network links. Set of resources provided by a particular

aggregate is called sliver.

User interacts with aggregates thru GENI's federated authorities called *clearinghouses*. Commonly used clearinghouses include the GPO Clearinghouse, Emulab and PlanetLab. Resources are generally links and nodes. Links are communications channels among nodes. Nodes are processors, routers, networks, or other configurable entities. All existing compute resource types (raw-pc, openVZ, Xen) are supported in stitched slivers.

A project organizes research in GENI, containing both people and their experiments. Project lead is the top authority for the project. Project lead manages adding people to project and experiments. Only people in the project have read and/or write privileges to experiments and resources.

Project and associated experiments are associated with a slice. A slice is the context for a particular set of experiments, and contains reserved resources and the set of GENI users who are entitled to act on those resources.

Security: Insufficient information. Various security improvements proposals …

GENI has unique features making security considerations unique (for now)

- Scale
- Interconnection with outside world
- Policy and procedural issues

Each GENI rack has a dataplane switch, which carries the traffic from experimenters in their GENI experiments. This switch is connected to the GENI network core, but also to a local network (e.g. a campus network) at each GENI rack site. This raises the possibility that traffic from GENI experiments could flow from the rack onto a site's regular network.

When site admins detect unwanted traffic flowing onto their network, they can block that traffic. Some ways to do that:

- Simple ACLs on the local device(s) that the rack dataplane switch connects to (in advance or configured in response to an incident to block unexpected traffic).
- A separate network firewall device in the path between the switch and the site network. Since the rack dataplane switches are intended to operate at gigabit-plus speeds, sites should take care to ensure that such a firewall device can handle that level of throughput.
- Configuration on the dataplane switch itself. The dataplane switch should not generally be configured by site admins to enforce local security policies proactively.

Configuration of resource approval - The rack software stack includes GENI aggregate managers that grant experimenters access to the rack's resources, including the dataplane switch.

Worth exploring – SDN routing and OpenFlow experimentation support within GENI.

A.6 NICT JAPAN STARBED^3 / JGN2+

INTRODUCTION

Primary domain(s)	IT Networking
High Level Description	StarBed^3 supports Internet research and experimentation using large scale simulation on a physical plant of 1000 computers interconnected by a 200 Gb/s internal backbone. External access to Japan's WIDE and JGN2+ networks is available.
Maturity level	The testbed is mature and supports users.

LOGISTICS

Website URL	http://starbed.nict.go.jp/en/ and http://www.jgn.nict.go.jp/jgn2plus_archive/english/about_us/nw.html
Owning Organization(s)	National Institute of Information and Communications Technology (NICT Japan)
Funding Source(s)	National Institute of Information and Communications Technology (NICT Japan)

USAGE

Research problems best used for	Large Scale internet simulations
Usage Restrictions	Permission of the operators. Access seems open.
Access Method	Remote access to dedicated computers and switches. Allocation of computers to users is done manually. Automated tools allow a researcher to allocate and configure their machines and switches programmatically.

INFRASTRUCTURE NATURE

A physical facility that provides physical computational and networking resources to researchers.
Up to 1000 physical computers that can be used to simulate or emulate thousands of computers.

INFRASTRUCTURE OPERATION

Researchers are allocated physical computers and switches manually and then can configure and use them programmatically. There are tools to support carrying out experiments that accept experiment specifications at a high level.

EXPERIMENT LIFECYCLE SUPPORT

StarBed^3 supports creation of experimental environments as well as programmatically carrying out experimental procedures. No features for experiment design. Archiving is somewhat ad hoc.

EXAMPLE EXPERIMENTS THAT USED THIS INFRASTRUCTURE

Strategic Technology to Prevent the Leaking of Information carried out by Hitachi and NEC. 1,020 virtual machines and 30 traffic generators. Traced router groups and P2P traffic and similar internet communications, such as e-mail, the web and etc., transmitted worldwide.
Television Conferencing System Technology carried out by the Panasonic Corporation. Used 100 physical computers to study multi-point HD video distribution.
IT-Keys Security Human Resource Training conducted personnel security training in 20+ isolated network environment containing 242 total computers.
Cloud Computing Competition carried out demonstrations and verification of a cloud computing competition. Participants created and demonstrated their applications in isolated network environments spread across 553 computers.

A.7 MIT LINCOLN LABORATORY (MIT-LL)

INTRODUCTION

Primary domain(s)	Information Systems and IT Networking
High Level Description	A comprehensive suite of cyber system assessment tools, including: range automation; network, host, and user emulation for IT systems; actuation of specialized embedded systems; complex experiment scenario modeling and execution; low-observable physical host instrumentation; attack framework.
Maturity level	Tools were either developed or extended for use in a medium scale, long running USCYBERCOM exercise. As such, they appear stable and ready for use elsewhere.

LOGISTICS

Website URL	https://www.ll.mit.edu/mission/cybersec/CSA/CSA-projects.html
Owning Organization(s)	MIT Lincoln Laboratory (MIT-LL)*
Funding Source(s)	Various U.S. Federal Government contracts

* SimSpace Corporation was recently formed to transition portions of the MIT-LL cyber system assessment tools to external use. As of the date of this publication, the details of which tools will be released and terms of release were still under consideration by the U.S. Government.

USAGE

Research problems best used for	Cybersecurity of distributed IT systems and networks that require a large testbed. Also possibly useful for experiments that include some range of cyber physical assets.
Usage Restrictions	The tools are available to U.S. Government and its contractors. Access must be requested.
Access Method	Install in your own facility. Some tools are available at the National Cyber Range (NCR) and may be used there, for those who have NCR access.

INFRASTRUCTURE NATURE

The MIT-LL suite of cyber system assessment tools include the following:

- Range Build-out and Automation. The MIT-LL tools use Visio with CCER extensions for range specification. Their ALIVE Range Automation software uses the range specification to set up the testbed for use.
- Lincoln Adaptable Real-time Information Assurance Test-bed (LARIAT) Traffic Generation. LARIAT provides network, host, and benign user emulation for IT systems at scale. It also provides Internet emulation. Desktop hosts supported include Windows, OS-X, and Linux. Applications supported include web, email, chat, and document editing.

It also supports Air Force tactical systems. LARIAT 9 now supports smartphones and tablets. LARIAT provides range situational awareness.

- Lincoln Lab Attack Framework (LLAF). LLAF automates existing open-source attack tools and exploits to emulate common attacks representative of Level 1 & 2 threat actors. LLAF is the malicious equivalent of LARIAT. It supports network reconnaissance, spear-phishing, and botnets.
- Goal-Oriented Scenario Modeling Robots (GOSMR) allows the user to model and execute complex experiment scenarios. GOSMR is an artificial intelligence (AI) framework that emulates human behavior. Unlike traditional user emulations that employ simple Markov chain models of human behavior, GOSMR models human behavior using techniques from research in economics and behavioral decision making. Each GOSMR robot agent has specified goals it must satisfy using human-like behaviors. E.g., GOSMR robots emulate the human tendency to procrastinate and perform at the "last minute".
- K0ALA Off-host Actuation. K0ALA is used to emulate users interacting with the consoles for specialized embedded systems. It can be used to automated systems ranging from military weapons systems to Xboxes. K0ALA is an off board device that interacts with the target system by intercepting and interpreting the video display and injecting mouse and keyboard events.
- LO-PHI Advanced Sensors. LO-PHI is a suite of instrumentation that can read transactions from buses and pull data structures directly out of system memory during test execution. LO-PHI introduces minimal detectable artifacts.

INFRASTRUCTURE OPERATION

See above.

EXPERIMENT LIFECYCLE SUPPORT

This software is used during experiment setup and execution.

EXAMPLE EXPERIMENTS THAT USED OR COULD USE THIS INFRASTRUCTURE

Blue team training. The MIT-LL tool suite can be used to set up and run training scenarios to teach and improve the skills of network and system defenders.

Red on blue exercises. The MIT-LL tool suite can be used to set up and run red on blue exercises as part of a larger experiment.

Defensive technique assessment. The MIT-LL tool suite can be used to set up and run evaluations of network defense techniques being researched.

A.8 TEST RESOURCE MANAGEMENT CENTER (TRMC) NATIONAL CYBER RANGE (NCR)

INTRODUCTION

Primary domain(s)	Information Systems and IT Networking
High Level Description	The National Cyber Range (NCR) is a facility that provides cybersecurity test and evaluation and training services to the US Government. It has a large, reconfigurable compute capability, enabling it to emulate DoD and other IT based environments for testing purposes. The facility has many advanced capabilities, but at present, these capabilities are not available for use by the much broader research community, which includes non-cleared, non-government funded researchers in private industry and academia. As of this report date, only some of the individual, unclassified tools used are available for use outside the NCR, whether via open source or on a fee basis.
Maturity level	The NCR is relatively mature and has been operating since 2012.

LOGISTICS

Website URL	https://www.tena-sda.org/display/intro/Home Registration is required.
Owning Organization(s)	U.S. Department of Defense (DoD) Test Resource Management Center (TRMC)
Funding Source(s)	The NCR is funded by the DoD TRMC.

USAGE

Research Problems Best Used For	Analyzing operations and security of information technologies under various conditions; questions of scalability; assessing system resiliency; examining system-of-systems.
Usage Restrictions	The NCR is a closed environment and is available only to U.S. Government organizations. Access and use must be coordinated through the NCR Director.
Access Method	Tests may be conducted on site or remotely via the Joint Information Operations Range (JIOR). Plans are being made to enable use through the Joint Mission Environment Test Capability (JMETC) Multiple Independent Levels of Security Network (JMN).

INFRASTRUCTURE NATURE

The NCR is a full service cyber testing facility, comprised of a secure facility, a security architecture, hardware, integrated tools for cyber testing, and a multi-disciplinary staff.

Base NCR capabilities:

- Support for all major operating systems and for IT services (e.g., email, web, domain controllers), network operations (e.g., routers, switches), and cybersecurity tools (e.g.,

intrusion detectors).

- Rapid configuration of complex, realistic network topologies and IT environment representations.
- Scales up to 40,000 virtual nodes of varying complexity and realism.
- Suite of data sensors and visualization tools.
- Unconstrained use of malware in tests.
- Supports use of physical wireless devices in tests, through use of a Faraday Cage.
- Simultaneously execute up to four independent tests using network isolation technology.
- On site expertise in offensive and defensive cybersecurity testing to assist in test design.

In addition to the above, the NCR is accredited by the Defense Intelligence Agency (DIA) and can operate at levels up to TS//SI-G/TK/HCS-P//SAR. It can:

- Simultaneously execute tests at different classification levels.
- Sanitize facility assets after tests for reuse in other tests at different classification levels.

INFRASTRUCTURE OPERATION

Use must be coordinated by the NCR Director and assisted by NCR staff members, to include resource allocation, custom device and software integration, data collection, and scheduling. Testbed build out is performed by NCR staff, who also conduct the tests.

Data collected during tests belongs to the customer and is protected per customer specification.

At test conclusion, a report is compiled and provided to the customer. All assets are then sanitized and returned to the pool of available resources for other customer use.

EXPERIMENT LIFECYCLE SUPPORT

Provides full experiment life cycle support.

EXAMPLE EXPERIMENTS THAT USED OR COULD USE THIS INFRASTRUCTURE

Wide variety of cybersecurity test types, including R&D testing, cybersecurity product evaluations, architectural analysis and assessments, and developmental and operational test and evaluation (DT&E/OT&E) in support of DoD acquisition programs.

Cyber Mission Forces training and exercise type events, target emulations, mission rehearsals, and malware and forensic analysis activities.

A.9 GEORGE MASON UNIVERSITY (GMU) OCTANE

INTRODUCTION

Primary domain(s)	Transportation (Automotive)
High Level Description	The George Mason University (GMU) Open Car Testbed And Network Experiments (OCTANE) provides a software package and a hardware framework for the reverse engineering and testing of automotive networks. OCTANE provides a platform for security research and teaching by replicating the interactions between the hardware components and control software of the systems so that the user can focus on the security aspects of the automotive network instead of the tool configuration and setup. (1)(2) It helps the analyst quickly identify and store all network packets, messages, and electronic control unit (ECU) IDs.
Maturity level	Moderately mature. The software is available to the research community under open source license.

(1) http://www.cs.gmu.edu/~mccoy/papers/CSET-GMU-OCTANE-paper.pdf

(2) https://www.usenix.org/conference/cset13/workshop-program/presentation/everett

LOGISTICS

Website URL	http://octane.gmu.edu/
Owning Organization(s)	George Mason University (GMU)
Funding Source(s)	National Science Foundation (NSF)

USAGE

Research problems best used for	Automotive network security
Usage Restrictions	Openly available for download via Octane web site.
Access Method	Physical/in person at GMU or via software download to set up one's own lab. Remote access to the GMU lab is not presently available but is a future goal.

INFRASTRUCTURE NATURE

Base, extensible open source Windows-based software package (C#) that one can use to connect to a vehicle's onboard network. Windows was chosen because a lot of drivers for cars are Windows only drivers while only a few have Linux drivers. Note: focus to date has been on the car's infotainment system.
Supports one vehicle at a time. There is no limitation on the make/model. You can plug into any car with a supported cable. (See below for supported cables)

INFRASTRUCTURE OPERATION

Plug into a vehicle's CAN network using supported cable connected to a Windows laptop. Use graphical user interface (GUI) to monitor network packets as the car operates. Simply do something with the car (e.g., lock doors) and watch the network traffic. Packets are decoded automatically to identify messages, commands, and electronic control unit (ECU) IDs. Commands are encoded using XML and can either be created via ASCII text editor or via the GUI and then injected into the message stream via the GUI.

Note: there is no car simulator. The system is plugged directly into a car's CAN network. There is presently no protection against overwriting the software on the car's computer systems.

OCTANE presently supports ECOM (3) and KVASER 5-700 cables (4). Other unsupported cables include AVR-CAN (5) and Intrepid (6). If a new cable emerges, support would have to be added for that cable. Can plug into any CAN network (7). In the future, support can be added for a LIN adapter or flexRay adapter.

Some vehicles are running Windows, some Linux, and some are moving to IOS. Most vehicles have a mix of operating systems. The actual engine control is usually Windows CE or a real-time OS. OCTANE does not presently have OS fingerprinting support, although it may in the future.

(3) ECOM cables are used mostly by researchers. CANCapture sells an ECOM cable @ $200. Software is not needed.

(4) KVASER is used mostly in Europe. Software is not needed, only the cable. See Kvaser.com.

(5) The AVR-CAN cable has one side serial, the other is COMM or COMM to USB. Support for this cable may be provided eventually, but there are presently no plans for such support.

(6) Intrepid cables are on the list of cables to support in the future.

(7) A CAN bus is a message-based protocol, designed specifically for automotive applications but is now also used in other areas such as aerospace, maritime, industrial automation and medical equipment. The CAN network is not TCP/IP on internal vehicle. The infotainment system has an IP stack that talks to external interfaces and will talk TCP/IP back to infrastructure such as On-STAR.

EXPERIMENT LIFECYCLE SUPPORT

Can provide network traffic monitoring and data collection. To monitor traffic, just plug in to a supported port with a supported cable.

EXAMPLE EXPERIMENTS THAT USED OR COULD USE THIS INFRASTRUCTURE

Can be used to monitor the car network while testing research solutions in areas such as:

- Firewalls (could possibly use an AVR-CAN to create a firewall or IDS as it has a CPU onboard)
- IDS
- Packet encryption
- ECU authentication
- ECU security

NOTE: OCTANE cannot support the debugging of timing issues on the CAN network. To do that, one would need to connect direct to the CAN interface and work in real-time. This interface is not powerful but could be used to prototype solutions. It has a JTAG interface.

A.10 OPEN NETWORKING LAB (ON.LAB) SOFTWARE DEFINED NETWORKING (SDN) TESTBED

INTRODUCTION

Primary domain(s)	IT Networking
High Level Description	The Open Networking Lab (ON.Lab) is virtual infrastructure and tools for research in software defined networking (SDN)
Maturity level	Unknown

LOGISTICS

Website URL	http://onlab.us/
Owning Organization(s)	ON.Lab
Funding Source(s)	Consortium (vendor) partners

USAGE

Research problems best used for	Industrial Strength solution for SDN "open source" codebase useful to service providers, telecoms, vendors. This testbed is actually purpose-built for testing the software product (a decentralized logically controller OS) for a multi-vendor SDN network. Overarching goal and usage pattern: to establish that it is feasible and possible to have 3rd parties run apps, Service Providers run infrastructure, all using open SDN technology from multiple vendors. Some applications of SDN environments, e.g. path computation, SDN-BGP AS gateway interconnection, etc. The applications are those that network operators, at-scale, would run on their own SDN infrastructure. Billing, accounting and resource accounting applications are also anticipated.
Usage Restrictions	Right now, contributors to open source codebase working with ON.Labs Future: possibly open (but not anticipated date or plan.) Internet2 test environment: open to the world.
Access Method	Internet (but vendors bring their physical switches to ON.Lab)

INFRASTRUCTURE NATURE

The ON.Lab testbed consists of virtual infrastructure for hosting VMs plus half-dozen vendors switches (vendors = 6) (switches = 12)
Control Plane centric experiments -- many controllers on VMs, reaching out to switch ports only when needed. Interoperability testing with real switches. Emulated switch environment for use instead of real switches. Sliced capability for slicing "real switches" -- work in progress

ONS demo 1, logical controller distributed over 8 controller nodes, 200-300 switches (not a problem to emulate 1000).

INFRASTRUCTURE OPERATION

Generally speaking, how does "it" work? E.g., a simulation/emulation, all physical nodes, etc.

Run VMs. Use scripts to configure software representing controllers and emulated switches. Use web/API to manipulate experimental environment.

EXPERIMENT LIFECYCLE SUPPORT

What parts of the experiment lifecycle does "it" support? (e.g., design, configuration, execution, data collection, etc.) What is the level of automation? (e.g., full vs. human assistant)

Tool - automated test harness (testON) -- written by Paxterra. Structured environment for organizing and running tests.

Open source tools to do continuous integration and automated testing.

Human driven manual testing to diagnose issues as they arise.

EXAMPLE EXPERIMENTS THAT USED THIS INFRASTRUCTURE

ONS demo to demonstrate resilience of open source OS (scaling, failure recovery, etc.)

New Abstraction for SDN Applications to make it easier to program the network.

Exercising and demonstrating SDN OS feature set; identifying future features of interest. ON.Lab follows an agile methodology where customer and partner demand drives additional of relevant features to their platform.

A.11 INTEL LABS OPEN CIRRUS

INTRODUCTION

Primary domain(s)	Cloud Computing
High Level Description	This cluster is designed to support academic research efforts, particularly research in cloud computing either funded by, or in collaboration with, Intel. Open Cirrus aims to address this problem by providing a single testbed of heterogeneous distributed data centers for systems, applications, services, and open source development research. The project is a joint initiative sponsored by Hewlett-Packard (HP), Intel, and Yahoo! The current (April 2010) testbed is composed of 10 sites in North America, Europe, and Asia. Each site consists of a cluster with at least 1,000 cores and associated storage.
Maturity level	Mature and fully operational

LOGISTICS

Website URL	http://opencirrus.intel-research.net/
Owning Organization(s)	Intel Labs
Funding Source(s)	

USAGE

Research problems best used for	Cloud security
Usage Restrictions	Need sponsor at Intel Labs
Access Method	Internet with approved account (**ssh bigdata.intel-research.net**)

UNDERLYING TECHNOLOGY DEPENDENCIES

Hadoop	**Apache Hadoop** is an open-source software framework for storage and large-scale processing of datasets on clusters of commodity hardware.
Maui & Torque	Scheduler and parallel batch system
Tashi	Manages virtual machines across a cluster
Ganglia	Monitoring system of clusters and grids; with GUI
Hadoop distributed file system (HDFS)	Cluster storage
Zoni	Physical machine allocation (still used? Mentioned in 2010 interview) to access “bare metal”

CYBERSECURITY EXPERIMENTATION OF THE FUTURE (CEF)

INFRASTRUCTURE NATURE

Multiple clusters of resources at various sites

INFRASTRUCTURE OPERATION

Unknown

EXPERIMENT LIFECYCLE SUPPORT

From bigdata, ssh to tashi. On this machine, cd into /usr/local/tashi to interact with Tashi. You can create, destroy, migrate, suspend, and resume virtual machines. Currently, images are stored on an NFS server - merkabah:/export/tashi/images.

EXAMPLE EXPERIMENTS THAT USED THIS INFRASTRUCTURE

From http://opencirrus.intel-research.net/projects/ (see below):

Not captured are (but explained on above web site):

- Food Recognizing Interactive Electronic Nutrition Display (FRIEND)
- Autolab - Autograding handin service for the cloud
- Register Allocation Deconstructed
- Big Data Cluster Management

Log-Based Architectures

A.12 RUTGERS ORBIT

INTRODUCTION

Primary domain(s)	IT Networking (Wireless)
High Level Description	The Open-Access Research Testbed For Next-Generation Wireless Networks (ORBIT) is a two-tier laboratory emulator/field trial network testbed designed to achieve reproducible experimentation, while also supporting evaluation of protocols and applications in real-world settings.
Maturity level	~11 years; actively maintained; they need new NSF grants (~every 4 years) to maintain the testbed by proposing extensions but this approach is not satisfactory

LOGISTICS

Website URL	http://www.winlab.rutgers.edu/docs/focus/ORBIT.html http://www.orbit-lab.org http://mytestbed.net/projects/omf/wiki/OMF_Main_Page
Owning Organization(s)	The project is a collaborative effort between several university research groups in the NY/NJ region: Rutgers, Columbia, and Princeton, along with industrial partners Lucent Bell Labs, IBM Research and Thomson. ORBIT is being developed and operated by WINLAB, Rutgers University
Funding Source(s)	National Science Foundation (NSF)

USAGE

Research problems best used for	The goal was to develop a large-scale wireless network testbed to facilitate a broad range of experimental research on next-generation protocols and application concepts. The ORBIT system consists of an indoor "radio grid emulator" for large-scale reproducible experiments, and an outdoor "field trial system" for subsequent real-world evaluations. A 400-node radio grid emulator has been set up in a dedicated 5000 sq-ft facility located at the WINLAB Tech Center building in North Brunswick in ~2004-05. The ORBIT radio grid was released for general use by the research community in Oct 2005, and has served over 500 research groups worldwide conducting a variety of experiments including mobile ad hoc networks, dynamic spectrum coordination, network virtualization, wireless security and vehicular networking. The testbed also serves as a proof-of-concept prototyping platform for wireless aspects of the NSF Global Environment for Network Innovation (GENI) future Internet infrastructure. An outdoor field trial network has also been set up with ORBIT nodes deployed at the WINLAB Tech Center and Busch campus.
Usage Restrictions	As ORBIT's name (Open-Access Research Testbed for Next-Generation Wireless Networks) says, almost all research or educational uses are appropriate and encouraged. These include use by universities, industrial

	research labs, and both US and non-US institutions. ORBIT users are organized in groups and user account authorization is handled by the group principal investigator (PI). PI is typically a faculty member or a senior/permanent staff member at the institution and is the person that is accountable for group members behavior while using the testbed. In order to create a new group, PI needs to fill in the new group registration form. Once the group is approved (typically less than a couple of business days), the PI will receive a group activation email. This will also establish the organization in ORBIT accounting system and allow new users to select that group while filling in the user registration form. The PI is solely responsible for approving new group user accounts and can delegate this role to any of the users in the group through the group management in the control panel. With some provisions, use for product development and evaluation by companies is also acceptable.
Access Method	On-site and remote (Internet portal); although field tests are almost exclusively on-site

INFRASTRUCTURE NATURE

Physical facility with 400 nodes with 2-10 wireless interfaces each in a large, shielded room

A small number of sandboxes (10-15) with 2 to 11 nodes (or more?) catering to specific "domains," e.g., OpenFlow, one with shielded wireless to create custom topologies reliably; used for testing before going to the large grid

OMF (ORBIT Management Framework) – a general-purpose experimentation framework being now maintained by NICTA (Australia); still used for ORBIT but also in other testbeds in Australia, Europe and the U.S.

ORBIT Measurement Framework & Library (OML)

Automatic scheduler for single-use experiment slots (30min – 2h)

Initial draft of Audit Experiment … Language to facilitate archiving experiments for reproducibility (data, metadata, versions etc.) but inherent conflict with implicit trust involved; NSF does not allow ORBIT to capture experimenter's information

Now 13 other U.S. locations with ~4 nodes or more each; connected via GENI

400 nodes maximum with more than 1500 wireless interfaces. Duration often minutes but experimentation slots are assigned as 30min – 2h. Experimenters can request subsequent 2h slots but not guaranteed.

INFRASTRUCTURE OPERATION

Generally this overall flow:

- Code development and testing (e.g., using Emulab, ns-2 or ns-3, or other means)
- Deploy on sandboxes
- Deploy on 400-node indoor grid
- Deploy in the field (outdoor)
- (Optional) Deploy at various ORBIT sites connected via GENI

EXPERIMENT LIFECYCLE SUPPORT

OML for measurements but users still need to define what to measure and post-process their

data; some guidance on capturing experimental setup as metadata but no automation or tools for this; provides fully automated experiment runs in automatically scheduled time slot(s) but many experimenters also use the testbed interactively to debug and run

EXAMPLE EXPERIMENTS THAT USED THIS INFRASTRUCTURE

Just finished hosting DARPA Spectrum Challenge (May 2014)
High density wireless – of current interest also in commercial domain

A.13 UNIVERSITY OF BUFFALO PHONELAB

INTRODUCTION

Primary domain(s)	Communications (Android Smart Phones)
High Level Description	PhoneLab is a public programmable Android testbed designed to simplify large-scale smartphone experiments. PhoneLab aims to address some of the common barriers that make experimentation at scale challenging—attracting participants and collecting data. PhoneLab participants receive discounted service in exchange for their participation in your experiments.
Maturity level	Operational with hundreds of participants (launched summer 2012)

LOGISTICS

Website URL	http://www.phone-lab.org/
Owning Organization(s)	University at Buffalo, NY State
Funding Source(s)	Sprint, National Science Foundation (NSF)

USAGE

Research problems best used for	Smart phone applications
Usage Restrictions	Sing-up online for experimenter status but must be affiliated with a university, school, or other suitable institution. Experiment subject privacy is protected. Experiments must be approved by an Institutional Review Board (IRB), a group of UB faculty members whose role is to safeguard the rights and welfare of research volunteers.
Access Method	Internet with approved account

INFRASTRUCTURE NATURE

Managing currently 288 participants that get a free smart phone at heavily discounted $45/month rate (unlimited talk, messaging, data); participants commit in 6 months intervals. In turn, participants agree to use this phone as their primary mobile device and participate in experiments; they choose which ones. The platform OS is custom built at PhoneLab and contains a Conductor app that manages the log collection and interaction with server.
100's of nodes; experiments run over months (up to a year) but experimenters don't know when and how many participants will join their experiment.

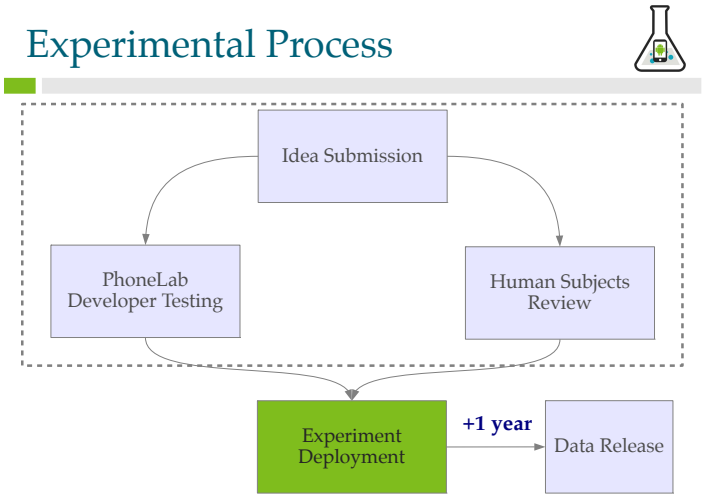
INFRASTRUCTURE OPERATION

While running the experiment, log data is first captured on the smartphone and then transmitted to PhoneLab's servers when the participant charges the phone. Experimenters can

then download the log data for their analysis. Experiments can work at app-level (easy deployment via Play Store) or at platform-level (distributed per over-the-air updates of the Android platform).

EXPERIMENT LIFECYCLE SUPPORT

As PhoneLab involves smartphone users ("participants") by design, the experiment lifecycle contains humans in all stages. The following slide shows the general flow. First, the experimenter submits an idea and obtains feedback from PhoneLab. In rare cases, the experiment is not promoted to the next step (e.g., improper use of the lab by commercial entities). Next, the IRB approval can be done in parallel to the testing done at the lab. Once both steps have finished, the experiment can be deployed; typically for up to 1 year (or however long the IRB has approved.)



The deployment phase is pull-based meaning that the experiment (either app-level via Google Play Store or a platform-level one integrated into an over-the-air update) is offered to participants and they select whether to run it. If app-level, the Play Store offers the standard ways of consent and alerts about what the app does and what data is collected.

Once participants are running the code, log data is collected on the device, then pushed to the PhoneLab server once the device is charging (so as not to influence mobile, "normal" usage)

Next year's upgrades will instrument the base OS image deployed on the participant's phones to collect a host of log information already (battery usage, network hand-off information etc.)

EXAMPLE EXPERIMENTS THAT USED THIS INFRASTRUCTURE

PocketParker

- Helping you find a parking spot

PhoneLab Usage Monitor

- Improving battery life
- Building a better Dropbox
- Monitoring campus wireless networks
- Umbrella experimental providing data for several studies:
  - Energy management and charging patterns
  - File usage

○ Wi-Fi access patterns
Flu propagation – Determining disease propagation through human contact
Characterizing privacy threats (University of Michigan) – Collections information about app usage patterns – Data collected will identify how frequently-used apps can leak information about users "Users tend to invoke a small set of apps frequently, and the rest of the apps with very low frequency, app sessions are short (80% less than 10 minutes). More than 98% of the app sessions happen when the users are visiting the same place (not while on the move)."
Improving mobile network access (University of Massachusetts, Amherst) – Logs last "network location" every 30 minutes – Aims to "further our understanding of the interplay of network mobility and geographic mobility" – Help design next-generation of the Internet built around mobility

A.14 IOWA STATE POWERCYBER

INTRODUCTION

Primary domain(s)	Energy (Electric Power Smart Grid)
High Level Description	The Iowa State University PowerCyber testbed provides a realistic electric grid control infrastructure based on a combination of physical, simulated, and emulated components. The testbed integrates industry standard control software, communication protocols, and field devices combined with power system simulators to provide an accurate representation of cyber-physical grid interdependencies. The testbed provides numerous cyber-security and power system research capabilities http://powercyber.ece.iastate.edu/publications/journal_testbed_2013.pdf
Maturity level	Operational

LOGISTICS

Website URL	http://powercyber.ece.iastate.edu/powercyber.html
Owning Organization(s)	Iowa State University
Funding Source(s)	NATIONAL SCIENCE FOUNDATION (NSF)

USAGE

Research problems best used for	Vulnerability assessment of field devices (relays, IEDs PMUs), protocols and automation software. System impact analysis, evaluation of risk mitigation and countermeasures, advanced persistent threat modeling. Attack-defense evaluations. Education and training.
Usage Restrictions	None
Access Method	Web-based remote access. Experiment templates provided for cyber-security scenarios.

INFRASTRUCTURE NATURE

PowerCyber is a power grid testbed with a focus on automation system vulnerability/security assessment, system impact studies, and attack-defense evaluations for real-world monitoring, protection, and control systems. The infrastructure provides access to a popular industry-grade SCADA environment from Siemens (PowerTG + Scicam) along with an array of protection and automation IED's from multiple vendors. The testbed has two means by which an emulated wide-area network may be provided. 1) ISEAGE is integrated into the physical environment itself and 2) DETERLab at USC-ISI has been successfully federated with PowerCyber as a lightweight (desktop) client. Multiple physical simulators are provided; on the electromagnetic side there is an RTDS and Opal-RT installation and on the electromechanical side PowerFactory is available. In terms of cybersecurity assets, the testbed has a variety of security tools (WireShark, nmap, Nessus) and the tools associated with ISEAGE including (D)DoS and Botnets.

The infrastructure currently has capability for experimentation up to ~40-bus power system, and the cyber system (substation automation systems) can scale into the hundreds using ISEAGE emulation. Efforts are currently underway to upgrade the system to simulate 100-bus power system in real-time, with necessary software/device upgrade.

INFRASTRUCTURE OPERATION

Physical assets such as IED's are interconnected over switches capable of VLAN isolation. Systems software (SCADA) is hosted on designated nodes within the environment and may be accessed via Remote Desktop. Using ISEAGE or DETERLab wide-area experiments using IEDs and professional SCADA software may be achieved.

EXPERIMENT LIFECYCLE SUPPORT

What parts of the experiment lifecycle does "it" support? (e.g., design, configuration, execution, data collection, etc.) What is the level of automation? (e.g., full vs. human assistant)

The testbed supports basic design and configuration of experiments through tools such as PowerFactory, RSCAD and Opal-RT Simulink extensions for the power side of things. ISEAGE provides a mechanism for cyber infrastructure description. Configuration, automation, execution and data collection are limited to point solutions for individual technologies.

EXAMPLE EXPERIMENTS THAT USED THIS INFRASTRUCTURE

Vulnerability Assessment on SCADA system – discovered several unknown vulnerabilities in Siemens SCADA system and "responsibly" disclosed to them to the vendor; two of these vulnerabilities are part of ICS CERT vulnerability repository, the discoverers are given credit. Reference: "ICSA-12-102-05—Siemens Scalance S Multiple Security Vulnerabilities", April 11, 2012.

Attack-Defense Experiment Demo at SmartAmerica Expo – Effects of coordinated attack – composed of data integrity attack and DOS attack – on Remedial Action Scheme (RAS), a wide-area protection scheme, implemented and demonstrated over the federated PowerCyber + DETERLab testbed infrastructure. In addition, the effectiveness of a perimeter-based DoS defense to mitigate the above attack scenario has been demonstrated ("live demo") at the SmartAmerica Expo held in Washington, DC, on June 11, 2014.

A.15 PACIFIC NORTHWEST NATIONAL LABORATORY (PNNL) POWERNET

INTRODUCTION

Primary domain(s)	Energy (Electric Power Smart Grid)
High Level Description	The power networking, equipment, and technology (powerNET) testbed provides power system and cyber equipment in a sandbox environment where current environments can be modeled for T&E or new architectures, algorithms, and controls can be explored and analyzed. PowerNET combines simulation, virtualization, emulation, and real cyber-physical equipment in one testbed to enable high fidelity and large scale experimentation for researching the effectiveness of electric power CPS applications and their cyber security implications.
Maturity level	Operational and can support internal and external projects.

LOGISTICS

Website URL	http://gridoptics.pnnl.gov/powernet/
Owning Organization(s)	Pacific Northwest National Laboratory (PNNL)
Funding Source(s)	Department of Energy (DoE), Department of Homeland Security (DHS), Private Industry

USAGE

Research problems best used for	Cyber-security in electric power CPS. New CPS architectures and communication paradigms. Demand response research. Operator training. Education. Validation and verification. Technology assessment and prototyping. Simulation and modeling. Demonstration.
Usage Restrictions	Requires approval to use.
Access Method	Remotely accessible and accessible by external researchers with approval.

INFRASTRUCTURE NATURE

PowerNET is a combination of IEDs (Relays, PMU's, and Smart Meters etc.), Simulators (Opal-RT, GridLab-D, PowerWorld ect.) and systems software (PDC's, HMI, EMS systems), infrastructure and network orchestration (Openstack, DETER). The ICS equipment, together with an IT infrastructure ties it all together in a cohesive research infrastructure. The interfaces to the system are geared toward both researchers and education (operator training, "hands-on" CPS training).
PowerNET can support 2 power substation environments with real equipment for high fidelity experimentation. For virtualized components the testbed can support 1000's of nodes for prolonged experiments. When researching large scale, simulation capabilities can be leveraged to scale to 10,000's of devices. All three capabilities can be merged to create research environments with user desired scale and fidelity. PowerNET also has federation capability that

has been demonstrated with research community partners like DeterLab at USC-ISI and the TCPIG testbed at UIUC.

INFRASTRUCTURE OPERATION

PowerNET provides an environment where simulators, IEDs and systems software are interconnected by an emulated wide-area networking environment (federated as needed). The aforementioned components are brought together with a series of operational servers and databases that collectively provide for coordinated inter-component functionality.

EXPERIMENT LIFECYCLE SUPPORT

The PowerNET testbed provides a fairly comprehensive set of experimentation lifecycle tools and capabilities. The design of the testbed includes database servers for experimental data warehousing and experiment scenario persistence, an orchestration server for experiment lifecycle automation, a device database for storing virtualized and physical device information, a network intrusion detection system, a simulation server environment and a user interface server for visualization and user interaction. By combining these assets in a systematic way PNNL provides extensive support experiment lifecycle.

EXAMPLE EXPERIMENTS THAT USED THIS INFRASTRUCTURE

Previous research topics include: distributed state estimation, GridLAB-D simulation use cases, DOE CEDS Digital Ants, quantifying deterministic behavior or SCADA network communication.
Conformance and interoperability testing of IEC 62351.
Hands on training with multiple academic and government institutions.
Multiple private party evaluations of network security applications in CPS environments, including Defense Industrial Base, and Silicon Valley software companies.

A.16 SKAION TRAFFIC GENERATOR

INTRODUCTION

Primary domain(s)	Information Systems and IT Networks
High Level Description	Generates realistic host based activity and network traffic at multiple scales and fidelities.
Maturity level	Seems mature but the customer base is very small. It has been used to support 10-15 large-scale tests. The Army is running it in Aberdeen. The 346 Test Squadron was running it in their range at end of the DARPA Dynamic Quarantine of Worms (DQW) project, but that was an earlier version. It is available in the National Cyber Range (NCR), but it is unclear whether it is being used there.

LOGISTICS

Website URL	www.skaion.com
Owning Organization(s)	Skaion
Funding Source(s)	Various U.S. Government contracts

USAGE

Research problems best used for	Any research project that requires realistic network traffic with limited protocols or host based activity.
Usage Restrictions	Currently must purchase software license with support. Typical cost is $20K for license that includes installation, support, and a small amount of time to integrate with custom applications. Skaion is looking into licensing opportunities with other companies.
Access Method	Install in your own facility.

INFRASTRUCTURE NATURE

The Skaion Traffic Generator is a software package consisting of three different traffic generators. Each provides a different level of scale and fidelity. The three traffic generators can be used together (mixed/matched) as desired.

- Console user
  - Realism: high fidelity
    - Looks like a real user
    - Does not run inside the host, so there is no footprint running on the host itself. Instead, controls the host the way the human does – interacting with the virtual keyboard, mouse, and screen.
    - Generates legit TCP/HTTP traffic
    - Distribution of traffic is based on what they got from AFRL
  - Scale: low scale

        - 100's implemented as one user per host
        - Each user is on a different virtual host (QEMU, VMWare Workstation+)
        - Can run with a user on a physical host if VNC is installed (including VNC on a dongle)
        - Has been used to emulate 100's of users
- Network traffic generator
    - Realism: medium fidelity
        - Works somewhat like the DETER threads.
        - Uses some real clients: web (http/https), email, ssh
        - Semi-realistic behavior -- will fetch a web page, sit for a while, then fetch a different page.
    - Scale: medium scale
        - 100's of users on virtual IPs on each host.
- Custom web servers with bot-clients
    - Goal is to generate the most realism possible at high scale
    - Realism: low fidelity
        - Generates valid http and file sizes are right, but timing may be off
        - Easy to tell it is not humans
    - Scale: high scale
        - High speed – looks like 100Ks users
        - Can saturate 100gb network with valid protocols
    - Meant to build the haystack of benign traffic in which to hide the interesting things

It comes with a fixed set of computer applications.

The design is modular, supporting two types of plug-ins:

- Application: An unofficial Python API exists that could be made public. The core components that move the mouse and look for patterns in the display exist and can be used by plug-in developers. To build an application plug-in, one only needs to build a Python app, put it into the applications folder, and modify a configuration file to automatically load the application.
- User behavior models (for console user traffic generator only): these modules are the "brains" or decision code that determines the actions of emulated console users. The Skaion tool has some tightly controlled / scripted models and one that uses Markov chains to just do "stuff", indiscriminate of what is going on environmentally.

INFRASTRUCTURE OPERATION

Software is installed on the base host/hypervisor, which accesses the VMs and interacts with them by controlling the virtual keyboard and mouse. It reads the screen by grabbing the images and processing them in GIMP. This has the benefit of having no footprint directly on the host, so it does not interfere with any host based malware or defensive mechanisms.

EXPERIMENT LIFECYCLE SUPPORT

This software is used during experiment execution. It has also been used to install software in VMs, so it could be used to automate node setup and collection of node logs post execution.

EXAMPLE EXPERIMENTS THAT USED OR COULD USE THIS INFRASTRUCTURE

Cloud-based scenario support: They developed a web spreadsheet software service for cloud and have clients that interact with cloud applications. Any web-based applications are already supported – they handle Google Docs now. They can also interact with virtual machines in a cloud.

Note: APL is working on an environment controlled by OpenStack that automatically sets up cloud environment. Can rapidly reconfigure a network in this core environment. Not a cloud of 1000's of machines, but dozens. This is a concept now, but doesn't exist yet.

SCADA operator-based scenarios could be supported. There is a short development cycle to add support for specialized applications.

Mobile/Tablet platform-based scenarios could be supported but not currently with physical devices. Mobile operating systems that can run in a VM (e.g., Android) can be supported with a small amount of integration. There may be a challenge to deal with morphing screens as the edges are not crisp. They believe support for some touch screens could be added – specifically if expressible through a hypervisor or if running inside QEMU, which has a "USB Finger". They believe that if a physical mobile device can run a VNC server on it (either directly or via dongle), then they could support the actual hardware.

Mission impact analysis scenarios could be supported. Users can create their own user behavior model plug in that implements a workflow. It would need to encode the complexities of alternatives to achieve the same goals. Likely would need to use AI.

Possible areas for government investment:

- Sharable user models.
- Directory of the entire research infrastructure that is available.

A.17 DEPARTMENT OF ENERGY (DOE) TCIPG

INTRODUCTION

Primary domain(s)	Energy (Electric Power Smart Grid)
High Level Description	While the Trustworthy Cyber Infrastructure for the Power Grid (TCIPG) project is a larger effort and includes R&D of hardware, software, protocols etc. in the trustworthy smart grid domain, the combined simulation/testbed environment has been developed as part of the project. It aims to accurately portray a representation of the electric power grid and allow for the exploration of cutting edge research at varying levels of scale.
Maturity level	Fully functional and in active use

LOGISTICS

Website URL	http://tcipg.org/research_Testbed http://tcipg.org/research_Modeling-Methodologies https://github.com/ITI/
Owning Organization(s)	University of Illinois at Urbana-Champaign, Dartmouth College, Cornell University, the University of California at Davis, and Washington State University
Funding Source(s)	Department of Energy (DoE) Office of Electricity Delivery and Energy Reliability, with support from the Department of Homeland Security (DHS) Science and Technology Directorate. TCIPG is the successor to the TCIP Center, funded by the National Science Foundation (NSF) (2005 to 2010).

USAGE

Research problems best used for	– Span transmission, distribution & metering, distributed generation, and home automation and control, providing true end-to-end capabilities. – Provide foundational support for TCIPG projects. – Analyze research across varying fidelities and scales. – Serve as a national resource for experimental work in research and analysis of trustworthy power grid systems.
Usage Restrictions	Research oriented, or under active projects with Illinois or TCIPG
Access Method	Physical or remote (access controlled), no open public access

INFRASTRUCTURE NATURE

The TCIPG testbed infrastructure is an extremely large collection of technologies brought together under one roof. If you can name it, it is likely that the testbed has the ability to conduct research on or with it. These technologies have been brought together for a wide range of experiments. The nature of the infrastructure is a centralized location providing access to a large

number of technologies.
Experiments can scale to thousands of nodes, where a node is anything from an embedded controller to a server computer. In AMI, there is the ability to model hundreds of thousands of nodes or greater, limited only by the amount of computation resources.

INFRASTRUCTURE OPERATION

Power systems, IED's, Control System and EMS software as well as physical simulators of the electromechanical and electromagnetic nature are interconnected in a testbed environment that provides dynamic configurability and varied network dynamics (both real-hardware and simulated systems with hardware interfacing). Full support for the most commonly used power systems protocols with full test harnesses.

EXPERIMENT LIFECYCLE SUPPORT

Due to the vast array of technologies in play, experimental development experience may vary widely depending on what technologies are required to accomplish a particular set of experimental goals. A unified approach to the experimentation lifecycle is not available at this time, however, TCIPG has created automation mechanisms for some heavily used components that elevate experimentation automation above the level of the point solutions provided by a particular technology, greatly facilitating testbed usability.

EXAMPLE EXPERIMENTS THAT USED THIS INFRASTRUCTURE

TCIPG Research (http://tcipg.org/research)
Los Alamos National Laboratory (quantum crypto validation)
Purdue University (AMI research)
University of Massachusetts (power fingerprinting for malware analysis)
Idaho National Laboratory (network observation and analysis)
Various company engagements
Lots of tools released to aid others in experimentation

A.18 APPLIED COMMUNICATIONS SCIENCES (ACS) VIRTUAL AD HOC NETWORK (VAN)

INTRODUCTION

Primary domain(s)	IT Networking (Wired and Wireless)
High Level Description	VAN (Virtual Ad hoc Network) provides a hybrid emulation environment for large network experimentation. Nodes communicate via *simulated* wireless and wired networks, where the simulated network can be represented by models in QualNet, OPNET, or ns-2 (support for ns-3 was planned for August 2014). VAN is suitable for experimentation with military tactical environments where wireless radio networks, including mobile ad hoc networks, are deployed. VAN has been used to support multiple cyber security projects for experimentation and demonstration purposes.
Maturity level	VAN has been in existence since 2008 and is relatively mature.

LOGISTICS

Website URL	http://www.appcomsci.com/research/tools/virtual-ad-hoc-network-van
Owning Organization(s)	Applied Communication Sciences (ACS)
Funding Source(s)	Previously funded by the Office of the Secretary of Defense (OSD) and U.S. Army Communications Electronics Research, Development, and Engineering Center (CERDEC). Currently being supported entirely by IR&D funds.

USAGE

Research problems best used for	Cyber security testing involving mobile wireless networks or requiring high fidelity for large-scale testing.
Usage Restrictions	VAN is presently available upon request with support on a fee basis. ACS may release VAN as an open source product for free download at a future date.
Access Method	The VAN testbed is being set up to allow Internet access with approved accounts. After the project source code is made publicly available, people will be able to download VAN software and create their own VAN testbed instantiations.

INFRASTRUCTURE NATURE

VAN is a suite of software tools that requires accompanying hardware (general purpose computers) and interconnecting equipment to install and run VAN software. In addition, a software simulator such as QualNet, OPNET, ns-2, or ns-3 is required for use in conjunction with VAN.

The scale of experiments that VAN can support depends on the fidelity required and on the available resources. VAN can be used to set up a test network with thousands of nodes without loss of fidelity.

INFRASTRUCTURE OPERATION

The VAN testbed consists of virtual machines communicating across a simulated network. VAN can also be connected to physical nodes and other emulated networks.

EXPERIMENT LIFECYCLE SUPPORT

VAN includes support for experiment management by providing the following experiment operations: select, deploy, start, pause, save and withdraw. Users can modify network state as part of an experiment (e.g., move node, disable link).

VAN supports "what-if" analysis during experiment execution. It supports debugging and analyzing an experiment's execution through temporarily stopping the experiment clock to pause an experiment.

VAN includes near real-time experiment situation awareness and post-test playback of events. It also includes metrics archival and a set of experiment analysis tools.

EXAMPLE EXPERIMENTS THAT USED OR COULD USE THIS INFRASTRUCTURE

Testing of a distributed intrusion detection system for enterprise and military wireless networks.
Moving target defense testing.
Military mobile ad hoc network experiments.
Large-scale network experimentation, using a network topology approximating an ISP's network running eBGP between Autonomous Systems (AS) and iBGP within each AS.

B CEF STUDY GROUPS

B.1 STUDY GROUP 1 AGENDA

Study Group on Hard Problems for Cybersecurity Experimentation of the Future (CEF)

March 13-14, 2014

SRI International
1100 Wilson Blvd., Suite 2800, Arlington, Virginia 22209

A G E N D A

DAY ONE – THURSDAY, MARCH 13

Welcome and Introductions – *David Balenson (SRI International)*

Opening Remarks – *Kevin Thompson, Bryan Lyles, and Farnam Jahanian (NSF)*

CEF Overview – *David Balenson and Laura Tinnel (SRI International)*
***Goal**: Provide background on the goals and structure of the overall CEF effort*

PANEL: Thoughts on Hard Problems for Cybersecurity Experimentation
***Moderator**: Terry Benzel (USC-ISI), **Panelists**: John Baras (UMD), David Nicol (Illinois), Bill Scherlis (CMU), John Wroclawski (USC-ISI)*

Breakout Group Instructions – *David Balenson (SRI International)*

BREAKOUT GROUPS: Hard Problems for Cybersecurity Experimentation
***Goal**: Identify cybersecurity hard problems and future research that would be amenable to or benefit from experimentation*
- *What are emerging or future societal dependencies on cyber technology?*
- *What hard cybersecurity problems will emerge from future cyber technology?*
- *What kind of future research can address cybersecurity challenges of future technology?*
- *What cybersecurity hard problems and future research will be amenable to experimental science?*

***Group A: Systems/Software,** Facilitator: Brian Decleene (BAE Systems)*
***Group B: Networks**, Facilitator: Steven Schwab (USC-ISI)*
***Group C: Cyber-Physical Systems**: Facilitator: David Corman (NSF)*

Breakout Group Report-Outs – *Facilitators and Scribes*

DAY TWO – FRIDAY, MARCH 14

PANEL: Thoughts on Approaches to Cybersecurity Experimentation of the Future

***Moderator**: Laura Tinnel (SRI International), **Panelists**: David Corman, (NSF), Brian DeCleene (BAE Systems), Anup Ghosh (Invincea), Steve Schwab (USC-ISI)*

BREAKOUT GROUPS: Approaches to Cybersecurity Experimentation of the Future

***Goal**: Identify cybersecurity experimentation approaches and methodologies that can address cybersecurity hard problems and future research*

- *What is the role of experimentation in research?*
- *What approaches and methodologies will best advance the research?*
- *What is the spectrum of capabilities that are desirable?*
- *Identify important characteristics, such as fidelity, scalability, and repeatability*

***Group A: Systems/Software,** Facilitator: Brian Decleene (BAE Systems)*

***Group B: Networks**, Facilitator: Steven Schwab (USC-ISI)*

***Group C: Cyber-Physical Systems**: Facilitator: David Corman (NSF)*

Breakout Group Report-Outs – *Facilitators and Scribes*

Wrap-Up/Summary/Next Steps – *SRI International and USC-ISI*

B.2 STUDY GROUP 2 AGENDA

Study Group on Infrastructure Needs for Cybersecurity Experimentation of the Future (CEF)

May 7-8, 2014

USC Information Sciences Institute
4676 Admiralty Way, 11th Floor, Marina del Rey, CA 90292

A G E N D A

DAY ONE – WEDNESDAY, MAY 7

Welcome and Introductions – *David Balenson (SRI International)*

CEF Overview – *David Balenson (SRI International)*

Study Group 1 Results: Hard Problems for CEF – *Terry Benzel (USC-ISI), David Balenson (SRI International), and Laura Tinnel (SRI International)*

PANEL: Thoughts on Experimentation Needs for CEF
***Moderator**: Terry Benzel (USC-ISI), **Panelists**: John Wroclawski (USC-ISI), Roy Maxion (CMU), Sam Weber, (CMU/SEI), Steve Corbató (U. Utah)*

Breakout Group Instructions

BREAKOUT GROUPS: Experimentation Needs for CEF
Goal: *Identify the types of experiments and experiment objectives needed to address research into the cybersecurity problems of the future*
- *What types of experimentation need to be conducted in the future?*
- *What are the goals and objectives of the experimentation?*
- *What experimental approaches and methodologies will best advance the research?*
- *Identify important characteristics, such as fidelity, scalability, and repeatability*

***Group A: Systems/Software,** Facilitator: Alefiya Hussain (USC-ISI)*
***Group B: Networks**, Facilitator: George Kesidis (Penn State)*
***Group C: Cyber-Physical Systems**: Facilitator: Zachary Tudor (SRI International)*

Breakout Group Report-Outs

DAY TWO – THURSDAY, MAY 8

PANEL: Thoughts on Infrastructure Needs for CEF

***Moderator**: Laura Tinnel (SRI International), **Panelists**: Thomas Edgar (PNNL), Damon McCoy (GMU), Phil Porras (SRI International), Lee Rossey (MIT-LL)*

BREAKOUT GROUPS: Infrastructure Needs for CEF

Goal: *Identify infrastructure and services needed to support the experiment types and objectives defined earlier*

- *What general capabilities (hardware, software, connectivity, etc.) are needed to support different types of experimentation?*
- *Identify specific tools needed to support experimentation*
- *Identify domain-specific needs*
- *Identify important characteristics, such as real world vs. emulated; centralized vs. distributed; independent vs. embedded*

***Group A: Systems/Software,** Facilitator: Alefiya Hussain (USC-ISI)*

***Group B: Networks**, Facilitator: George Kesidis (Penn State)*

***Group C: Cyber-Physical Systems**: Facilitator: Zachary Tudor (SRI International)*

Breakout Group Report-Outs

Wrap-Up/Summary/Next Steps

B.3 STUDY GROUP 3 AGENDA

Study Group on Capability Needs for Cybersecurity Experimentation of the Future (CEF)

June 18-19, 2014

SRI International
1100 Wilson Blvd., Suite 2800, Arlington, Virginia 22209

A G E N D A

DAY ONE – WEDNESDAY, JUNE 18

Welcome and Introductions – *David Balenson (SRI International)*

CEF Overview – *David Balenson (SRI International)*

Study Groups 1 and 2 Results: Hard Problems and Experimentation Needs for CEF – *Terry Benzel (USC-ISI)*

PANEL: Thoughts on Capability Needs for CEF
***Moderator**: Terry Benzel (USC-ISI), **Panelists**: Marshall Brinn (Raytheon BBN), Scott Lewandowski (The Wynstone Group), Steve Schwab (USC-ISI)*

Breakout Group Instructions

BREAKOUT GROUPS: Capability Needs for CEF
Goal: *Identify infrastructure and services needed to support the experiment types and objectives expected in the future*

- *What general capabilities (hardware, software, connectivity, etc.) are needed to support different types of experimentation?*
- *Identify specific tools needed to support experimentation*
- *Identify domain-specific needs*
- *Identify important characteristics, such as real world vs. emulated; centralized vs. distributed; independent vs. embedded*

***Group A: Experiment Design,** Facilitator: Ritu Chadha (ACS)*
***Group B: Experiment Realization,** Facilitator: Sami Saydjari (CDA)*
***Group C: Experiment Analysis,** Facilitator: Dale Johnson (MITRE)*

Breakout Group Report-Outs

DAY TWO – THURSDAY, JUNE 19

Overview of Existing Experimentation Infrastructure – *Laura Tinnel (SRI International)*

BREAKOUT GROUPS: Prioritized Capability Needs for CEF

Goal: *Identify and prioritize missing cybersecurity experimentation capabilities needed over the coming years, for research challenges spanning multiple sectors*

- *What existing "cornerstone" capabilities can we build on?*
- *What are the gaps between current and needed capabilities?*
- *What are the overlaps and differences in needs by domain?*
- *What are the near-term, medium-term, and long-term capabilities that need to be addressed?*

Group A: Experiment Design, *Facilitator: Ritu Chadha (ACS)*
Group B: Experiment Realization, *Facilitator: Sami Saydjari (CDA)*
Group C: Experiment Analysis, *Facilitator: Dale Johnson (MITRE)*

Breakout Group Report-Outs

Wrap-Up/Summary/Next Steps

B.4 STUDY GROUP PARTICIPANTS

A total of 75 individual from 50 organizations participated in the three CEF study groups:

Aaron Johnson, Naval Research Laboratory
Alefiya Hussain, USC Information Sciences Institute
Andre Weimerskirch, University of Michigan Transportation Research Institute
Angelos Keromytis, National Science Foundation
Anil Somayaji, Carlton University
Anita Nikolich, National Science Foundation
Anthony Joseph, University of California, Berkeley
Anup Ghosh, Invincea
Benjamin Edwards, University of New Mexico
Bill Scherlis, Carnegie Mellon University
Brad Martin, National Security Agency
Brandon Schlinker, University of Southern California
Brian DeCleene, BAE Systems
Bryan Lyles, National Science Foundation
Charles Palmer, IBM Research
Dale Johnson, The MITRE Corporation
Damon McCoy, George Mason University
Dan Cantu, Sandia National Laboratory
David Balenson, SRI International
David Corman, National Science Foundation
David Nicol, University of Illinois at Urbana-Champaign
Donna Dodson, National Institute of Technology and Standards
Elaine Shi, University of Maryland
Ersin Usun, Xerox PARC
Ethan Katz-Bassett, University of Southern California
George Kesidis, Pennsylvania State University
Gianluca Stringhini, University of California, Santa Barbara
Gideon Juve, USC Information Sciences Institute
Grant Wagner, NSA Research
Gregg Schudel, Cisco Systems, Inc.
Herb Lin, National Academies
James St. Pierre, National Institute of Technology and Standards
Jean Camp, Indiana University
John Baras, University of Maryland
John McHugh, University of North Carolina and Redjack
John Sebes, TrustTheVote Project
John Wroclawski, USC Information Sciences Institute
Josiah Dykstra, National Security Agency
Kevin Butler, University of Oregon
Kevin Sullivan, University of Virginia
Kevin Thompson, National Science Foundation
Laura Tinnel, SRI International
Lee Rossey, MIT Lincoln Laboratory
Luke Berndt, Department of Homeland Security Science & Technology Directorate
Manimaran Govindarasu, Iowa State University

Marshall Brinn, Raytheon BBN
Mary Denz, Air Force Research Laboratory
Matthew Elder, Symantec
Maverick Woo, Carnegie Mellon University
Micah Sherr , Georgetown University
Miles McQueen, Idaho National Laboratory
Paul Boynton, National Institute of Technology and Standards
Phil Porras, SRI International
Ritu Chadha, Applied Communication Sciences
Roy Maxion, Carnegie Mellon University
Ryan Goodfellow, Washington State
Sam Weber, Carnegie Mellon University Software Engineering Institute
Sami Saydjari, Cyber Defense Agency (CDA)
Sandy Clark, University of Pennsylvania
Scott Lewandowski, The Wynstone Group
Sean Peisert, University of California, Davis
Sonia Fahmy, Purdue University
Steve Corbato, University of Utah
Steve Schwab, USC Information Sciences Institute
Ted Faber, USC Information Sciences Institute
Terry Benzel, USC Information Sciences Institute
Terry Champion, Skaion Corporation
Thanassis Avgerinos, Carnegie Mellon University
Thomas Carroll, Pacific Northwest National Laboratory
Thomas Edgar, Pacific Northwest National Laboratory
Tim Yardley, University of Illinois at Urbana-Champaign
Vaibhav Garg, Drexel University
Vincent Urias, Sandia National Laboratories
Von Welch, Indiana University
Zach Tudor, SRI International

C CEF ADVISORY GROUP

An advisory group comprised of government, industry, and academic senior leaders to help inform and guide the CEF study. The advisory group was comprised of the following individuals:

Dr. Michael Donald Bailey
Associate Professor
University of Illinois Urbana-Champaign (formerly University of Michigan)

Dr. Steve Corbato
Director of Cyberinfrastructure Strategic Initiatives
Office of Information Technology, University of Utah

Dr. Steven E. King
Deputy Director for Cyber Security
Information Systems and Cyber Security Directorate, Office of the Assistant Secretary of Defense for Research and Engineering

Mr. John Lowry
Senior Technical Director
Raytheon BBN

Dr. Douglas Maughan
Cyber Security Division Director
Homeland Security Advanced Project Agency
Department of Homeland Security Science and Technology Directorate

Dr. William H. Sanders
Department Head and Donald Biggar Willett Professor of Engineering
Department of Electrical and Computer Engineering,
University of Illinois Urbana-Champaign

Dr. Patrick Gerard Traynor
Associate Professor
Department of Computer and Information Science and Engineering
University of Florida

SRI International®

USC Viterbi
School of Engineering
Information Sciences Institute

This material is based upon work supported by the National Science Foundation under Grant No. ACI-1346277 and ACI-1346285. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.