This is an Implementation proposal. The proposal has no significant environmental or cultural impacts.

1 Executive Summary

We propose to build innovative, transformative research infrastructure (RI) for cybersecurity and privacy (CS&P) experimentation, titled "Mid-scale RI-1 (M1:IP): SPHERE – Security and Privacy Heterogeneous Environment for Reproducible Experimentation". As described in Section 4, CS&P threats impact our daily lives, our national infrastructures and our industry. CS&P are of critical importance to all other sciences, to protect confidentiality, integrity and availability of scientific discoveries and scientific infrastructures. Unfortunately, CS&P research today is piecemeal, opportunistic, simplified and not reproducible, mostly due to use of private data on private infrastructure, which limits collaboration, cross-pollination and vertical progress in these scientific fields.

The CS&P research community needs a common, rich, representative research infrastructure, which meets the needs across all members of the community, and facilitates reproducible science. Such infrastructure is needed to transform CS&P research from piecemeal, opportunistic and simplified, to highly integrated and reproducible. We describe the community need in Subsection 4.1, with evidence including: *NSF's Ten Big Ideas* [1], *White House Multi-agency R&D Priorities for 2023* [2], *FY 2023 Federal Cybersecurity R&D Strategic Plan Implementation Roadmap* [3], *Presidential Policy Directive 21 (PPD-21)*, *Cybersecurity Experimentation of the Future (CEF) study* [4] and *CEF 2022 Workshop [5]*, *National Academies' Report on Reproducibility and Replicability in Science* [6], and *Cybersecurity Artifacts Workshop* [7]. We also discuss why existing RIs do not meet unique experimentation needs of CS&P research.

Required component	Proposal section
a. Implementation or Design	First page of description
b. Environmental or cultural impact	First page of description
c. Results from prior NSF support	Section 9
d. Scientific justification	Section 2
e. Tangible benefits	Section 2
f. Preliminary activities	Subsection 5.7
g. Implementation plan	Subsection 5.8
h. Operation and utilization plan	Subsection 5.9
i. Broader impacts	Section 3
j. Institutional commitment to diversity	Section 7
k. Divestment	Section 8
l. Foreign collaborators	None

Table 1: Pointers to required components.

The SPHERE RI is illustrated in Figure 1 and further described in Section 5. To facilitate study of varied threat sources and potential mitigation strategies, the RI will offer rich, abundant, and diverse hardware resources, which would meet the experimental needs of 90% of researchers today [8]. Many CS&P researchers study phenomena that interact closely with network topology, protocols and actors – SPHERE will meet the field's unique needs by offering a dedicated, user-configurable network substrate. CS&P experiments further may include generation of harmful traffic, taking live measurements from the real Internet, running human user studies, and even interacting with malicious Internet actors. To support these different research needs, and protect the Internet, SPHERE will provide safe network security policies.

All SPHERE nodes will be accessible via a single user interface. To meet the needs of various classes of users, SPHERE will provide six user portals: MAN (manual) - for exploratory research, JUP (Jupyter) – for mature research, GUI – for novice users, EDU – for use in education, AEC – for artifact evaluation committees and HUM – for human user studies. Users will be able to access all portals from the user interface, and obtain a consistent view of their experiments, while being able to switch between portals as their needs evolve. Research use cases are highlighted in red in Figure 1, and example research supported by SPHERE is summarized in Subsection 5.6. SPHERE will facilitate reproducible science by building a streamlined process, standards, and incentives for the community to develop representative (realistic) experimentation environments (REEs –Subsection 5.2) and offering built-in RI supports and community engagement process for artifact sharing and reuse (Subsection 5.4).

Table 1 provides further pointers to required elements of the proposal.

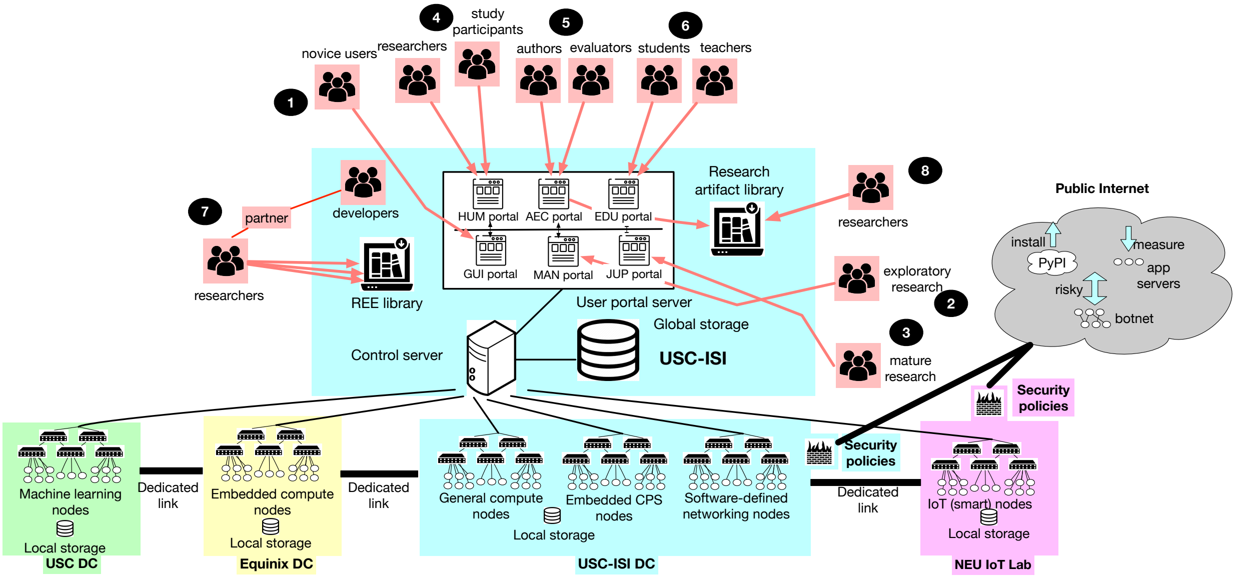


Figure 1: SPHERE architecture – different colors indicate different physical hosting locations. All equipment will be accessible through a single user interface, with tabs to switch between multiple portals, to support different use scenarios. Use scenarios are shown in red.

2 Intellectual Merit

Scientific Justification: The SPHERE research infrastructure will offer six novel research capabilities to meet CS&P community needs: (1) access to heterogeneous, contemporary hardware, which is currently being targeted by cyber attacks or used to deploy flexible cyber defenses, (2) ability to combine multiple types of hardware resources into a single experiment, via a configurable network substrate and protected by safe security policies, (3) six unique user portals, accessed via a single user interface, supporting different classes of users, and lowering barriers to entry, (4) a streamlined process, standards and incentives to help CS&P community develop and deploy high-quality, representative experimentation environments on SPHERE, (5) built-in, full support for reproducibility to help users produce high-quality, reusable research artifacts, (6) a dedicated portal and community engagement plan for artifact evaluation committees, to build a pipeline, feeding research artifacts into SPHERE. See Subsection 4.2 for prior work comparison.

The SPHERE research infrastructure will *transform* CS&P research, from piecemeal and opportunistic to highly integrated, by unifying the research community's experimentation efforts on a common, rich, highly usable infrastructure. Through its unique hardware resources, SPHERE will facilitate novel research on emerging technologies, and will ultimately lead to timely securing of new technologies and prevention of attacks, allowing defenders to outpace attackers. SPHERE's representative experimentation environments, built-in support for reproducibility, and alliances with artifact evaluation committees will enable vertical progress in the science of CS&P. By increasing the pace of innovation and improving the success and sophistication of CS&P research products, SPHERE will significantly advance scientific discovery and the Nation's research capabilities in CS&P. SPHERE will further enable richer, more realistic experimentation through a combination of its unique hardware resources, support for human-user and Internet measurement experiments, and representative experimentation environments. This will lead to more deployable research products and faster technology transition, thus improving security of all other sciences by protecting their data, workflows, and infrastructures. Accelerating the pace of innovation and breaking down barriers to research and education progress will enable the U.S. to lead the world with sophisticated cybersecurity and privacy solutions in these areas of both national and global importance.

Tangible Benefits: SPHERE will provide a *common, rich, representative research infrastructure, which meets the needs across all members of the community, and facilitates reproducible science*. This need is described in Subsection 4.1 with evidence including: *NSF's Ten Big Ideas* [1], multiple federal roadmaps and policy directives [2, 3, 9] and multiple community reports and workshops [4, 5, 7, 6]. SPHERE will offer *common, rich research infrastructure* by providing a large number of devices in different hardware classes, connecting them by a dedicated, configurable network substrate and enabling interactive but safe experimentation via its security policies. SPHERE will *meet the needs across all members of the community* via: (1) its blend of heterogeneous hardware resources (projected to meet the needs of nearly 90% of experiments in current CS&P publications [8]), (2) its user portals aligned with needs of different classes of users, and (3) its unique support for human user studies, Internet measurement, and experiments including live Internet interactions. SPHERE will facilitate *reproducible science* by providing a process, a standard and incentives for the community to develop representative (realistic) experimentation environments (REEs), and by offering built-in support and a community engagement process for artifact sharing and reuse.

The SPHERE research infrastructure will facilitate novel, transformative research endeavors in CS&P: (1) fully reproducible research, where research artifacts hosted on SPHERE can be easily and reliably reused, (2) deployable research, whose products are developed and tested in representative environments, which reflect deployment targets and diversity of human user behaviors; such research can easily transition to commercial products, (3) integrated research, reflecting progress of the research community leading to more sophisticated and effective research products, (4) broadly-applicable research, whose products use online machine learning to adapt to changes in threats and deployment environments, and use software-defined networks and hardware for flexible deployment, (5) research on novel threats and defenses, enabled by specialized hardware hosted on SPHERE, such as trusted execution environments, various edge computing nodes, various IoT nodes, programmable logic controllers, etc. These novel scientific explorations will make U.S. researchers world leaders in CS&P science.

3 Broader Impacts

SPHERE will change the dynamics of today's reactive CS&P research, leading to effective, deployable research products and to a more secure, more private digital nation. Achieving this vision requires leading-edge CS&P research, conducted in a common, rich, representative, and usable research infrastructure (RI) that promotes reproducible science. SPHERE will move CS&P research and education from the piecemeal, opportunistic, simplified efforts of individual labs and educators, towards integrated and sophisticated research and education that reflects the contributions of many community members. The broader impacts of the proposed research infrastructure include a faster pace of innovation in cybersecurity and privacy fields, more mature solutions on the market meeting stakeholder needs, as well as impact of these solutions broadly across the U.S. scientific and economic communities.

Advances in cybersecurity and privacy are essential across a wide range of sciences and digital infrastructures (e.g., observatories, nuclear facilities, computational infrastructure, supply chains and business processes) to secure their workflows, datasets and infrastructure against cyber attacks and information leaks. CS&P advances will impact current and emerging technologies, leading to more secure online space for users in both everyday and critical services.

SPHERE will enable reproducible experimentation on shared hardware that is easily and remotely accessible by all U.S. researchers. This will democratize security and privacy research, and will especially benefit underserved researchers and students, enabling them to compete on an equal standing with those from top-tier institutions. Through its reproducibility services, SPHERE will make available many CS&P artifacts and thus fuel research publications.

SPHERE will further enrich and broaden participation in CS&P education and strengthen the U.S.

Threat	Relevant CS&P research
Theft of scientific data	Prevention and mitigation approaches for intrusions and data exfiltration
Loss of access to scientific infrastructures	Ransomware and denial of service mitigation
Impersonation of scientific servers and users	Authentication, anti-phishing, zero-trust architectures
Scientific data and code modification	Data integrity approaches
Sensitive (e.g., patient) data exposure	Secure multi-party computation, secure federated learning

Table 2: Relevance of cybersecurity and privacy research to other sciences.

workforce by offering a common platform and a rich set of education materials. This will enable students to experiment with and learn about the cutting edge threats and defenses in CS&P. Students will develop marketable, practical skills, thus making them highly competitive in today's global labor market. The SPHERE project will further directly involve underserved student populations in its development. Our training program will fund 20 interns per year (80 students total), recruited solely from underserved student communities, to participate in SPHERE's implementation.

We will make all SPHERE code and reports publicly available through GitLab repositories and our project Web site.

4 Relevance of Security and Privacy Research and Community Need

Over the past decade, and especially during the Covid-19 pandemic, both an individual's and society's essential functions (e.g., work, school, entertainment, social, financial, infrastructure, and governance) moved increasingly online. This sharply increased our nation's dependence on correct and reliable functioning of network and computing systems, and has led to increases in the frequency and impact of cybersecurity and privacy (CS&P) attacks. Recent years have seen unprecedented and record-breaking attacks, for example the Solar Winds supply-chain attack [10], which exposed confidential government data, and the Colonial Pipeline attack [11], which shut down our major gas pipeline for several days. Ransomware attacks more than tripled [12], DDoS attacks doubled [13], and data breaches increased by 70% [14]. Simply put, we now live in a world where cybersecurity and privacy are intrinsically intertwined with everything we do, and failures in these domains can have far-reaching monetary and national security impacts, and even jeopardize human lives. **Research progress in cybersecurity and privacy is thus of critical national importance**, to ensure safety of U.S. people, infrastructure and data.

Cybersecurity and privacy are crosscutting issues, affecting many scientific fields, as summarized in Table 2. Scientific infrastructure and individual researchers are increasingly becoming an appealing target for cyber threats that look to steal research data and hijack our research progress (*confidentiality threat*), to modify research data or impersonate scientific infrastructure (*integrity threat*), or to make scientific infrastructure unavailable for U.S. researchers (*availability threat*). During Covid-19 pandemic, China-sponsored cyberattacks brought down scientific infrastructure and tried to steal research data [15, 16, 17, 18]. Mid-2022, Iranian hackers tried to steal credentials from and infect computers of researchers in Middle Eastern affairs, nuclear security, and genome research [19]. In late 2022, Alma observatory in Chile came offline due to cyberattacks [20, 21, 22, 23], and Russian hackers impersonated login servers of U.S. nuclear facilities [24] to steal researchers' credentials. If obtained, these credentials could be used to access confidential data about U.S. nuclear capabilities. **Advancing cybersecurity and privacy research will help us secure our scientific cyber-infrastructure and ensure progress in many scientific disciplines**, enabling safe collaboration and secure communication between scientific groups, ensuring data confidentiality, integrity and availability, and facilitating distributed computations that preserve privacy. The cross-cutting nature of cybersecurity and privacy is also reflected in the NSF's funding model – the SaTC and CICI programs are crosscutting across multiple directorates and the NSF's large TrustedCI project helps secure large NSF facilities against attacks.

Cybersecurity and privacy are broad and multifaceted scientific fields, encompassing complex interactions between technology, individuals, social norms, and organizations. As technology progresses, attack surfaces also increase, moving from traditional computing systems to IoT devices, embedded devices and critical infrastructure – each of these environments poses unique challenges for cybersecurity and privacy research. CS&P's complexity, coupled with the fast pace of innovation (which increases attack surface) is not met by current research practices. Current research is opportunistic. It is performed piecemeal, using tools, datasets and equipment available to a given research lab, often in isolation from other related work and often on much simplified versions of larger, interdependent systems. New publications frequently fail to build on past research, nor do they quantitatively compare their findings with prior publications. In fact, such comparison is often prohibitively expensive, as it requires reproducing other researchers' work, and overcoming insurmountable obstacles: missing, incomplete or unusable research code or data, and lack of access to special hardware used in prior research. **The current isolated, simplified and piecemeal research practices** make defenders lag behind attackers, prevent less-resourced researchers from contributing to the field, and **undermine the U.S.'s ability to lead in the science of cybersecurity and privacy**. Security and privacy advances that are simplified, poorly evaluated, or misaligned with user priorities fail to secure our society against cyber attacks, exposing cyber-infrastructure in other scientific disciplines to attacks, and leaving our critical infrastructure vulnerable and our economy suffering heavy losses.

4.1 Community Need

CS&P researchers need **common, rich, representative research infrastructure, which meets the needs across all members of the community, and facilitates reproducible science** to move from *piecemeal, opportunistic research* to *pursuing integrated, sophisticated, community-encompassing research.*

Evidence of Community Need: The need for *common, rich RI in CS&P domain* is supported by the following reports and directives. *NSF's Ten Big Ideas* [1] calls for dynamic and flexible approaches to RI and for harnessing the data evolution by enabling crosscutting research, education, and advanced cyberinfrastructure. *White House Multi-agency R&D Priorities for 2023* [2] calls for development of cybersecurity research infrastructure. *FY 2023 Federal Cybersecurity R&D Strategic Plan Implementation Roadmap* [3] lists six priority areas, related to CS&P. *Presidential Policy Directive 21 (PPD-21)* defines 16 critical infrastructure sectors [9], which pose distinct challenges to CS&P, and therefore necessitate a comparably complex and heterogeneous RI. The *Cybersecurity Experimentation of the Future (CEF) study* [4] and *CEF 2022 workshop* [4] identified new directions for CS&P experimentation to include: (1) encompassing various domains of applicability and (2) interconnected RI with heterogeneous resources. *To quantify the need for heterogeneous resources, we conducted a survey of publications at top cybersecurity conferences in 2022* [8] and found that out of 656 publications, experimental needs of 35% can be met by general compute nodes, while the remaining 65% required more specialized hardware and software.

The need for *representative RI in CS&P* is supported by the *CEF study* [4] and *CEF 2022 workshop* [4], which call for (1) modeling the real world for scientifically sound experiments, (2) frameworks and building blocks for extensibility and usability, (3) representative environments, and (4) support for human user studies and live Internet traffic.

The need for *RI that facilitates reproducible science* is supported by several community reports. *National Academies' Report on Reproducibility and Replicability in Science* [6] outlined multiple unmet needs for reproducibility and replicability, and publications [25, 26] highlight that reproducibility is especially challenging for computer science and computer systems research. *Cybersecurity Artifacts Workshop* [7] identified problems around research artifact's findability and usability in CS&P, and researchers' need for infrastructure support for research artifact packaging and reuse. This need for RI support for reproducibility was echoed in the *CEF 2022 workshop* [5].

4.2 Existing Research Infrastructure and Efforts

We now discuss why the existing research infrastructures (RIs) do not meet the CS&P community need and summarize our points in Table 3.

Common, rich RI for CS&P must offer heterogeneous nodes, a configurable network substrate and it must support safe experimentation. DeterLab [27, 28] is the only public cyberinfrastructure specializing in supporting CS&P experimentation. It has been run by us for two decades and is approaching its end of funding. DeterLab currently only hosts general compute servers, which meets the needs of only 35% of published papers [8]. DeterLab offers a configurable network substrate and supports experimentation fully isolated from the Internet. It is thus unable to support CS&P research directions that require live Internet interactions. Other research infrastructures (Chameleon [29], CloudLab [30], POWDER [31], AERPAW [32], Cosmos [33], FABRIC [34], SAGE [35]) host resources for experimentation in a specific field of communication (wireless, mobile, programmable networks, machine learning, data-center networking, sensing). These infrastructures offer high-fidelity environments for experimentation in their domain of science, but lack the diversity of resources needed for the majority of CS&P experiments. No current public research infrastructure offers experimental access to nodes with trusted execution environments, a variety of embedded compute nodes, CPS nodes and IoT nodes, and there is limited support for programmable switches (CloudLab [30], FABRIC [34]). Most of the existing infrastructures offer only experimental nodes, while a few also offer configurable network substrate. None offer the flexible security policies needed by CS&P researchers. *SPHERE will host six hardware classes, all relevant to CS&P research, with multiple device types in each class, and multiple instances of each device type (Subsection 5.1). Nodes will be connected via a dedicated, configurable network substrate and experimenters will be able to choose between four safe security policies.*

Representative RI for CS&P must offer experimentation environments that resemble real-life deployments of CS&P technology. Since existing RIs focus on supporting research in other scientific domains, they do not offer representative experimentation environments for CS&P research. *SPHERE will build a streamlined process, standards, and incentives in the research community, to develop a set of representative experimentation environments for CS&P (Subsection 5.2).*

RI, which meets the needs across all members of the community must offer user interfaces aligned with different classes of users: novices vs experts, researchers vs students/teachers. Current RIs offer a limited set of user interfaces, which are cumbersome for novices and students, and are either geared towards exploratory research (command-line interface), or towards mature research (Jupyter notebooks). *SPHERE will offer six user portals, aligned with needs of different classes of users (Subsection 5.3).*

RI, which supports reproducible science. Existing RIs offer some support for experiment packaging to promote reproducibility, mostly via creating node images (disk, VM or Docker images). Node imaging is necessary, but not sufficient for reproducibility. It records the experiment setup, but it does not capture the experimental workflow, which occurs after the setup. Node images are further not portable to other hardware, nor to newer OS versions. A few RIs offer limited support for saving experimental workflows, mostly via Jupyter notebooks [36, 37]). This works well for computational workflows, which often execute the same actions on each experimental node and engage a small number of applications to process datasets. The Jupyter notebook approach, however, is not well-aligned with CS&P experiments. CS&P experiments often require execution of different actions on different nodes within a single experiment, where actions depend on each other, and many involve OS events in addition to application events. The complexity of CS&P experiments requires a unique approach to workflows, geared to support of action dependencies (Subsection 5.3). We have pioneered such approach in DEW – distributed experiment workflows [38], and we plan to extend it to SPHERE. A few existing RIs support research artifact sharing (DeterLab's shared artifacts [39], Chameleon's Trovi [40], CloudLab's profiles [41]) via RI-hosted libraries. *SPHERE will*

RI	Rich hw for CS&P	Safe ex.	Conf. net.	UIs	REEs for CS&P	Reproducibility	
						images	workflow
Chameleon	GC	no	limited	CLI, Jupyter	no	yes	yes
CloudLab	GC, SDN	no	yes	CLI, NS scripts	no	yes	no
POWDER	GC, SDR	no	yes	CLI	no	yes	no
AERPAW	GC, UAV	no	no	CLI, Google form	no	yes	no
Cosmos	SDR, RF, FPGA, SDN	no	yes	CLI	no	yes	no
FABRIC	GC, storage, SDN	no	yes	CLI, Jupyter	no	yes	yes
SAGE	Sensors, ML	no	no	Python extensions	no	yes	yes
SPHERE	GC, EC, ML, SDN, CPS, IoT	yes	yes	CLI, Jupyter, GUI AEC, EDU, HUM	yes	yes	yes

Table 3: Comparison of existing RIs with SPHERE. GC – general compute, EC – embedded compute, ML – machine learning, SDN – software-defined networking, SDR – software-defined radio, FPGA – field-programmable gate array, UAV – unmanned arial vehicle, RF – radio frequency, CLI – command-line interface, CPS – cyber-physical system, IoT – Internet of things, GUI – graphical user interface.

offer built-in reproducibility support for research artifact packaging, sharing and reuse, including logging user actions to ease cognitive burden of whole experiment packaging, experiment testing for portability, and artifact libraries (Subsection 5.4).

The CS&P research community has increasingly embraced the formalization of artifact submission and evaluation to improve the quality of research and increase vertical progress. Yet, only a few CS&P conferences today have artifact evaluation, more than half of the papers do not submit artifacts, and less than a quarter obtain the "results replicated" badge. *SPHERE will enhance and accelerate reproducibility efforts in CS&P, by building alliances with artifact evaluation committees (AECs), offering an AEC-specific portal, and directly supporting artifact evaluation and hosting on SPHERE.*

5 Details of the Proposed Research Infrastructure

In this section, we detail physical design and buildout plans in Subsection 5.1, representative experimentation environments in Subsection 5.2, user portals in Subsection 5.3, reproducibility support in Subsection 5.4 and engagement and outreach in Subsection 5.5.

5.1 Physical Infrastructure

Figure 1 illustrates the proposed SPHERE architecture, showing different classes of devices and using color to denote where each class will be hosted. SPHERE's equipment will be hosted at multiple *facilities*. Each facility hosts its own infrastructure for storage, experiment control, networking and security. Most SPHERE equipment will be hosted at USC-ISI datacenter, which incurs no separate hosting costs. Equipment with requirements beyond what USC-ISI DC can meet (e.g. large power supply, special racking requirements) will be hosted professionally at the USC datacenter (USC DC in the figure) and the Equinix datacenter (Equinix DC). Each datacenter offers a unique trade-off with regard to hosting cost, power and space restrictions. There is dedicated, 100 Gbps bandwidth between the three datacenters (USC-ISI, USC and Equinix). Northeastern University has new dedicated space that will host IoT nodes (NEU IoT Lab in the Figure).

While there is no dedicated Layer-2 connectivity between USC-ISI and NEU, there is Layer-3 connectivity via Internet 2. We will start by leveraging this Layer-3 connectivity to build experiments that span NEU IoT Lab and the rest of SPHERE. This approach meets the needs of the majority of users, providing reliable connectivity, and configurable bandwidth, delay and loss, with some variability due to mixing of experimental and Internet traffic. Over the course of the project, if our users need fully-dedicated Layer-2 connectivity between USC-ISI and NEU, we will pursue this option. The FABRIC project already has a presence in Los Angeles, CA (where USC-ISI is located) and Boston, MA (where NEU is located) and can provide a dedicated, configurable network between these two locations. To achieve fully dedicated Layer-2 connection between USC-ISI and NEU, we will work with NEU and USC-ISI's upstream Internet providers to explore use of AL2S links to connect NEU and USC-ISI to FABRIC ports in Los Angeles and Boston.

Table 4: Details of the proposed hardware and broad classes of CS&P research it supports

General compute nodes: 48 from DeterLab, 144 new nodes, with Intel TDX, ARM CCA/TrustZone, and AMD SEV **Research supported:** application, system and network security, measurement, human user studies, large-scale experiments, education, trustworthy computing
Machine learning nodes: 10 GPU-equipped servers **Research supported:** security with machine-learning in the loop
Cyber-physical nodes: 15 Rockwell Automation ControlLogix PLCs, I/O modules **Research supported:** critical infrastructure security
Embedded compute nodes: 600 from DCOMP, 312 new (Intel Atom, Intel Xeon D, ARM Cortex-A57, and NVIDIA Jetson NX Volta GPUs) **Research supported:** edge computing security, blockchain security, private computing, trustworthy edge computing, federated learning
IoT nodes: 500 IoT nodes (a variety of smart home, smart speaker, camera, doorbell, TV, appliance, medical, office, wearable, and miscellaneous devices) **Research supported:** IoT security, user privacy
Programmable nodes: 8 Tofino switches, 16 Xilinx Virtex-7 NetFPGA development boards (smartNICs) **Research supported:** dynamic (programmable) network security, SDN security

The devices we plan to purchase and integrate with SPHERE as *experimental nodes*, and the research that benefits from these are shown in Table 4. SPHERE will support most popular and relevant devices for CS&P research today. If CS&P research trends change in the future, new devices can be easily added by adding new installation and control scripts to the Merge software.

SPHERE user, control, storage and security infrastructure. *Approach:* We will build on our mature RI control software – Merge[42], which currently supports three research infrastructures at USC-ISI (DeterLab, DCOMP, Searchlight) with diverse hardware and hundreds of users. Merge software supports a resource commissioning process: resource owners declare models of the *capabilities* and *connectivity* of their resources to the management portal. These models provide information to automated installers that generate installation scripts for the various nodes/switches, using established scripting languages (e.g., Ignition [43], Zero-Touch Provisioning [44], Ansible [45]). Merge's design allows for extensible RI, where new node types can be quickly integrated, without major code redesign. Merge further offers outstanding performance for large-scale experiments, allocating and provisioning thousands of virtual machines within one minute. Merge installers and software are developed and tested using best practices, including repository-level unit testing as well as integration and chaos testing in virtualized test environments. In this task, we will purchase hardware and install appropriate software for Merge infrastructure, which includes a unified user portal server, implemented as a computational cluster for scalability and resilience, disk imaging and OS loading infrastructure, large storage local to each facility, global storage at the user portal server, user authentication and access control infrastructure, and security infrastructure. In the proposed work, we will further extend Merge's control software to support *instrumenting and managing non-standard devices*, such as IoT devices, CPS devices, GPU-enabled compute nodes and programmable switches and network cards. Further, we will extend Merge to support commissioning and driving IoT sensors and actuators, such as robot button pushers, voice synthesizers, microphones and cameras. *Benefits to the research community:* Supports the need for experimentation with heterogeneous devices in the CS&P community. Facilitates large-scale experimentation and supports extensible research infrastructure.

General compute nodes. *Approach:* We will absorb 48 newer general compute nodes from DeterLab (NSF Award #2016643) and purchase 144 high-density server class nodes, to support a mixture of virtualized and bare metal experimentation on different chip architectures. The nodes will be equipped with novel hardware for trusted execution, including Intel SGX, Intel TDX, ARM CCA/TrustZone, and AMD SEV. When virtualized, they will support nearly 10,000 virtual nodes. *Benefits to the research community:* Virtualization will enable sharing of physical nodes among multiple experiments, to support large-scale research experiments, and to use SPHERE in education. Bare-metal nodes will be useful to researchers who require high-fidelity measurements without virtualization overhead or trustworthy computing.

Machine-learning nodes. *Approach:* We will purchase 10 GPU-equipped server-class nodes and ensure high-bandwidth connectivity to mass storage resources for efficient access to large datasets. *Benefits to research community:* Machine learning is redefining boundaries of the possible in many areas of science, including security and privacy [46, 47]. CS&P experiments require data capture (e.g., from observations of system events or traffic) in real time and use of the data for ML-enabled, real-time attack mitigation.

Cyber-physical nodes. *Approach:* We will purchase 15 popular PLCs (e.g., Rockwell Automation ControlLogix), and the corresponding I/O modules that connect PLCs to the physical parts of an industrial control system, as well as the corresponding networking infrastructure. We will purchase licensed software to program PLCs, and use Rockwell Automation's virtual asset management to virtualize SCADA, human-machine interface and engineering workstations on our general compute nodes. The PLCs and I/O modules will interface with state-of-the-art digital twin modeling and simulation, such as Rockwell Automation's smart manufacturing digital twin modeling software [48] or open-source digital twin modeling software [49]. *Benefits to the research community:* These resources will provide a platform to test and experiment with safety-critical industrial control processes at all levels of the operational stack, in a controlled environment and via remote access. Researchers will study the physical impact of cyber vulnerabilities in real industrial development environments, without damaging expensive equipment and without the significant effort required for physical experiment configuration. The flexibility of the process simulation allows researchers to target various safety-critical domains.

Embedded compute nodes. *Approach.* We will absorb 600 Intel Atom-based Minnowboard nodes from DCOMP, and purchase 312 diverse embedded compute nodes (Intel Atom, Intel Xeon D, ARM Cortex-A57, and NVIDIA Jetson NX Volta GPUs), some of which will have Intel SGX or ARM Trustzone capabilities. *Benefits to research community.* A large number of diverse embedded compute nodes in the SPHERE RI will facilitate realistic, large-scale, edge-to-cloud experimentation (including other SPHERE resources) with distributed applications, including issues around device security, device mobility, data integrity, data confidentiality (privacy), trustworthy edge computing and computation fairness.

IoT nodes. *Approach:* We will purchase 500 IoT devices of various types (see Table 4). We will also purchase devices necessary for remote access to IoT nodes: *IoT sensors*, to remotely report the status of the devices, which includes cameras and microphones; *IoT actuators* — tools to remotely activate and interact with the IoT devices, which take non-standard input, such as: (1) robotic arms and button pushers to interact with touch-screens and other physical interfaces; (2) infrared transmitters to simulate movement for movement sensors and to control TV devices; (3) controlled speakers and voice synthesizers to interact with smart speakers and voice assistants; (4) other companion devices, such as smart phones, for devices controlled by a companion app. We will develop tools for capturing the network traffic produced by IoT devices and for streamlining automated experiments at scale. *Benefits to research community:* SPHERE's

IoT nodes will be the first to enable *repeatable, controlled, remote* IoT experimentation at scale, along with the ability to perform *physical interactions remotely*, thus facilitating reproducible IoT security and privacy research for many researchers in a common environment.

Programmable nodes. *Approach:* We will purchase 8 Tofino SDN switches, supporting P4, and 16 smart NICs (Xilinx Virtex-7 NetFPGA development boards). These devices will be remotely accessible and usable in custom experiment topologies, including any of the other SPHERE nodes. *Benefits to research community:* The ability to experiment with programmable switches and NetFPGAs in *CS&P scenarios* will result in products that offer high performance and are readily deployable in today's modern networks. These nodes can also be used to study security of SDN technologies.

5.2 Representative Experimentation Environments

In the CEF study [4] and the CEF 2022 workshop [5], researchers overwhelmingly expressed needs for representative (aka "realistic"), pre-configured experimentation environments (whole experiments, tools and datasets). It is difficult to create one broadly representative experimentation environment (REE), because representativeness depends on the research problem being studied in the experiment. For example, an environment that is representative for studying link flooding attacks will include a bottleneck topology, many attackers, some legitimate clients and servers, some legitimate application traffic and high-bandwidth attack traffic generators. In contrast, an environment representative for studying ransomware will include a sophisticated business network topology, many different devices, simulated or real users, and realistic user interactions with information stored on the devices. Due to the complexity of real-life scenarios, it is unlikely that a single research group can alone develop a REE for a broad range of study directions. A research group may develop a REE for one narrow research direction, and another group may improve on it. It is our vision that, over time, the entire research community engages with this task, building on each others' work and iteratively improving and refining experimentation environments, until they become highly sophisticated, representative and applicable to multiple study directions.

We will *build partnerships between domain experts and RI developers* to create and deploy a large set of REEs. In such a partnership, the researcher (domain expert) provides expertise about what is representative for their research direction, and the RI developer helps replicate such an environment on SPHERE. We will build a large set of REEs by: (1) building a seed set of REEs ourselves by porting products of our prior RI-developer/domain-expert partnerships to SPHERE, (2) soliciting applications by researchers for supplemental funding (included in our budget) to transform their research artifacts into REEs on SPHERE, (3) implementing safe security policies to support experiments that include live Internet interactions.

Developer-contributed REEs. USC-ISI developers have served as the test and evaluation teams for several DARPA-funded projects in CS&P: SAFER, EdgeCT, XD3 and DComp. During each program our team produced REEs that were used to evaluate the research products of multiple teams in common, challenging scenarios. These environments, which we will port to SPHERE, are summarized in Table 5.

Table 5: Developer-contributed REEs. DDoS – distributed denial-of-service, C – client, S – server, A – attacker

Anonymous comm. Tor network with realistic bw and delay per geolocation
Confidential comm. Red (secure enclave) / Black (public Internet) networks, multiple topologies
Distributed attacks. Internet-level AS topology with configurable link properties
Low-rate DDoS. Large-scale C/S topologies with realistic traffic mix
Extreme-rate DDoS. Large-scale C/S/A topologies with realistic network topology and trafic mix
CDN attacks. Realistic CDN topology and traffic mix

Community-contributed REEs. We will build partnerships between RI developers and domain experts (researchers), to add more REEs to SPHERE. Use case 7 in Figure 1 illustrates how REEs will be contributed by researchers, who partner with RI developers, and how REEs will be stored in a publicly accessible library. Each year, we will run multiple open solicitations for research artifact contributions from the research community. We will define criteria for artifact eligibility, such as the artifact being a part of top-tier publication and being cited or reused by others. Multiple discussions around reproducibility identified lack of funding as key blocker [7, 5, 6]. We will provide support to up to 10 research artifact authors per year (cost included in our budget for two months for a student), to improve their artifact and port it to SPHERE as an REE. Each author will be paired with a SPHERE developer for this task. Each artifact will have to pass a strict set of acceptance criteria before becoming an REE, including verification of usability by two independent researchers.

Internet Interaction Environments. CS&P experiments require security policies to protect other experiments on SPHERE and the Internet from harmful traffic. Yet some experiments also require interactions with the Internet, e.g., to support software installation, to incorporate metrics based on live Internet interactions, or to interact with a malicious actor over the Internet. In the SPHERE RI, we will develop four security policies: (1) default policy: blocking all traffic between the experiment and the Internet, (2) installation-specific policy: allowing traffic between the experiment and the user-specified, software installation portals (e.g., PyPI), (3) measurement-specific policy: allowing user-specified probe traffic between the experiment and user-specified Internet destinations; (4) open-but-cautious policy: allowing all communication between the experiment and the Internet, but applying special monitoring [50] to detect and drop potentially destructive, malicious or excessive traffic.

5.3 User Portals

Researchers have overwhelmingly expressed the need for intuitive, easy-to-use user interfaces. [4, 5]. In SPHERE, we will develop six user portals. All SPHERE nodes will be accessible via a single user interface, with tabs allowing the user to switch between different portals as their needs evolve. *Experiment state will persist across portals and remain consistent. The portals simply offer a different view of an experiment, a different way to interact with it, and sometimes an additional set of functionalities (e.g., access to education materials in EDU portal).* Each portal will support the entire experiment lifecycle – creation, resource allocation, running (workflow), monitoring, data collection, data post-processing, resource release, and experiment termination. By design, the portals will address the needs of diverse user classes and ease user onboarding. The three foundational portals support different levels of researcher experience and experiment maturity (GUI, MAN, JUP; use cases 1, 2, and 3 in Figure 1). Three additional portals each expose the foundational portals, but add new features to support specific use cases: the growing need for human-in-the-loop experimentation (HUM portal, Figure 1's use case 4), artifact evaluation by conferences and journals (AEC portal, Figure 1's use case 5), and use of SPHERE in education to support workforce training (EDU portal, Figure 1's use case 6).

The **GUI portal** will extend our prior work on DEW: Distributed Experiment Workflows (NSF Award #1835608) and is essential to support novice users. GUI portal will especially benefit students in underserved populations, who often lack experience with research infrastructure. The GUI will support drag-and-drop interactions to design an experiment's topology and workflow, to run the experiment and monitor its progress. These features already exist in DEW [38, 51], and will be extended to support new node types. Our GUI portal will further host a library of possible topologies and workflow actions, thus enabling fast, meaningful experiments by novice users. The **MAN portal** will support SSH access to experiment nodes' terminal application; thereby allowing for exploratory research, e.g., when a graduate student starts working on a new experiment and manually interacts with nodes, trying different design choices. The **JUP portal** will support experimentation scripts, encoded as Jupyter notebooks, allowing for mature research by experienced users, i.e., when exploring a range of experiment parameters, and running many times to ensure statistical significance of results.

CS&P experiment workflows often involve traffic and system events in intricate scenarios, with a given event triggered by the start or the completion of one or more prior events – we refer to this type of workflow as *a workflow with event dependencies or WWED*. Supporting WWEDs is imperative to ensure the validity of results in CS&P experiments, and it can shorten experiment time by aborting early on failure. Our GUI portal will inherit WWED support from DEW portal [38], which already has a graphical user interface, and underlying mechanisms to support WWED on DeterLab. Once we extend WWED support to all nodes in SPHERE, we will also add it to MAN and JUP portals. This will include: (1) prompting users to define if their consecutive actions on experimental nodes are dependent or not, and (2) wrapping user commands into WWED constructs to enforce dependencies. This is a unique capability, not available in today's RIs.

The **HUM portal** will allow for human-in-the-loop experiments. Researchers will design their experiment using the underlying MAN, JUP, and GUI portals, then use the HUM portal's unique functionality to expose one or more experimental nodes via an experiment-dedicated Web interface. The interface will serve as the means for human study participants to interact with the experiment, and the means by which researchers collect and annotate data. This combination of rich CS&P experimental infrastructure and human interaction is critical for research in how human users both adopt threat prevention/mitigation strategies and create vulnerabilities through their actions. *Note that SPHERE will simply provide an environment that can be used to let human study participants interact with experiments on SPHERE (e.g., test a new authentication scheme). It will be the responsibility of SPHERE users to recruit their study participants, obtain IRB approvals from their institutions, and manage any data from participants accordingly.*

The **AEC portal** will support artifact evaluation. The evaluation team will port a research artifact to SPHERE using the underlying MAN, JUP, and GUI portals. The AEC portal will further allow the team to take notes during evaluation (e.g., how they overcame a given software installation challenge), and exchange private messages with each other and anonymous messages with the artifact authors. The portal will capture the history of evaluation actions, using logging features from our artifact packaging services, which can provide valuable information about the artifact's initial usability and the time burden to reuse it.

Building on our pioneering work on using public testbeds in CS&P education (see: Section 6), the **EDU portal** will allow *instructors* to assign work to students, manage student accounts, monitor student work, and download assignments submitted by students. It will allow *students* to interact with the RI via the underlying GUI, MAN or JUP portals to complete their assignments, and submit completed work through EDU portal. We will further develop a set of teaching materials for SPHERE by porting existing CS&P teaching materials from DeterLab to SPHERE (previously used to educate ≈20,000 students from almost 150 institutions) and adding 2-3 new education modules *per node class* (10-15 modules total) that leverage IoT, CPS, embedded compute, ML and programmable nodes. We will further build guidelines and processes for other educators to contribute education modules to SPHERE. Broad educational opportunities are critical to build U.S. capacity in CS&P domain through a large, capable workforce, which has first-hand, practical experience in CS&P threats and defenses.

5.4 Reproducibility Support

We will develop built-in services for experiment packaging, sharing and reuse, which will help promote easy reuse of CS&P artifacts. CS&P researchers often experiment over a long time period (weeks to months), trying various manual and scripted actions to produce a result for publication. Due to a large number of actions and a long time period of experimentation, there is a high cognitive burden on researchers to identify the correct set of actions to be shared, so that others could reproduce their results. It is thus important to provide cognitive aids to users on our RI to help with this task.

User action logging. Because users often use a command-line interface to interact with their experiments during exploratory research, it is necessary to log user actions on the experiment nodes and in all user portals. For each action, we want to log user input, the action's output (partial output may be sufficient), action's time and location (e.g., which experimental node or portal the action was run on). The most complicated part of this task is logging user actions on nodes' command line interfaces. We will extend our prior work on command-line logs of user actions on general compute nodes [52], to support logging on all SPHERE devices. We will then develop ways to identify sequences of successful actions in logs and present them to users. Users will be able to select actions to package into a workflow for a given experiment. This is still a cognitive challenge to users, due to the sheer scale of actions, which could be on the order of several hundred per day of use. We will develop ways to compress and expand the user view, and to let users tag the actions during RI use, so they can easily identify them later.

Artifact packaging. The Cybersecurity Artifacts Workshop [7] identified missing or incomplete documentation as a significant cause of low artifact usability. Workshop participants suggested development of community standards for artifact packaging. In this task we will develop built-in packaging support for experimental artifacts (whole experiments, tools and datasets). This includes interactions with the researcher community and holding of meetings, workshops and asynchronous discussions to standardize artifact metadata, and writing up user guidelines and standard documents. When packaging an experiment, we will want to package its setup (topology, software and configurations on nodes), workflow, and input and output datasets. In addition to packaging of node software and configuration as images, we will further implement Packer-based packaging, to promote portability of experiments. The Packer tool [53] is a free, open-source tool for creating OS images for multiple platforms from a single source configuration. We will promote use of Packer for experiment setup, but we will also develop ways to identify setup actions (software installation and node OS and application configuration changes) from logs of user activity and transform them into Packer scripts. We will also package experiment workflows. We will provide ways to automatically test a packaged experiment for resilience (does it always run?), consistency of results (does it produce same outcomes within some margin of error?) and portability (can an experiment be reused by a different user on a different experiment?). Finally we will develop automated tools to test if a packaged artifact was properly documented, according to the community standards.

Artifact libraries. We will develop built-in support for libraries of experimental artifacts, to support both AEC-contributed and user-contributed artifacts. Artifacts in the library will be easily accessible from designated users and owners will be able to decide whether to keep them private, share them with selected users or all users of SPHERE. Artifacts will further be easy to reuse, thanks to standardized artifact metadata.

5.5 Outreach and Engagement

We propose extensive outreach and engagement activities to meet several goals: (1) learn about the needs of CS&P researchers, and adjust our infrastructure implementation plans as the project progresses, (2) obtain beta users for SPHERE, (3) involve students from populations underrepresented in computing in SPHERE's implementation, (4) promote wider sharing and reuse of experimental artifacts and (5) promote use of SPHERE in CS&P education.

Need-discovery workshops and surveys. We will run three need-discovery workshops and surveys per year. The workshops will be co-located with different CS&P conferences (e.g., NDSS, IEEE S&P, Usenix Security, ACM CCS, ACSAC), and will inform our project activities. Online surveys will also be distributed broadly to researchers via emails, taking care to target diverse research communities, such as networking, CPS, IoT, SDN, etc.

SPHERE RI promotion activities. We will develop tutorials and videos, showcasing both simple and complex experiment functionalities. We will also offer three virtual, live office hours per year. To popularize use of SPHERE in education, and broaden participation of underrepresented groups in this important field, we will work directly with instructors at community colleges, state universities, and minority-serving institutions. We already have connections to many such instructors via our prior DeterLab education work. We will further reach out to researchers at MSIs to promote SPHERE for their research. USC has in the past offered a joint cybersecurity course with Florida Agricultural and Mechanical University (FAMU), an HBCU. We will continue offering this course and include use of SPHERE in its syllabus.

Student training. We will offer paid, in-person, student training at USC-ISI over each summer during the project duration. Student interns will work with SPHERE developers on implementation tasks. We will recruit 20 students per year from community colleges, black and Hispanic-serving colleges and universities, and two-year colleges. Many of these organizations are already using DeterLab infrastructure in education, and we have strong ties to the course instructors, which we will use for recruitment. We will also broadly advertise this opportunity at minority-serving events, such as Tapia conference, SACNAS, Grace Hopper conference, etc.

Alliance with artifact evaluation committees. We plan to develop alliances with artifact evaluation committees (AECs) at CS&P venues, with the goal to support artifact evaluation on SPHERE infrastructure. The committees would advertise SPHERE and work with artifact authors to evaluate their work on SPHERE. Artifacts that are successfully reproduced will be preserved in our artifact libraries to be easily reused by future researchers (we envision a library per venue, plus a global SPHERE library with pointers to all shared artifacts, including those shared directly by users). *We have explicitly discussed this with the Program Committee Chairs and AEC Chairs for NDSS and ACSAC, who have agreed to consider SPHERE as a platform for artifact evaluation whenever possible, and to work with us to support long-term hosting of artifacts for the research community.*

5.6 Example Research Supported by SPHERE

In addition to the general research use cases, shown in red in Figure 1, we illustrate a few non-exhaustive examples of how SPHERE uniquely enables innovative research.

Studying industrial control system (ICS) security in a realistic environment. SPHERE will enable researchers to run cyber-physical security experiments on safety-critical ICS scenarios with high fidelity. As depicted in Figure 2, our CPS infrastructure will leverage standard programmable logic controllers (PLCs) and distributed control architecture, coupled with physical world simulations via digital twin software, to offer both high experimentation fidelity (cyberspace) and potential to create endless experimentation scenarios (physical space). SPHERE RI will provide researchers with a realistic environment to study attacks that exploit vulnerabilities in certain PLCs and their interactions, to evaluate impact of such attacks on physical functionality of a given ICS, and to build and evaluate attack mitigations. SPHERE RI will enable researchers to interface with actual operational data to model and conduct experiments on real-world system vulnerabilities using state-of-the-art ICS control technology. Promising mitigations, which perform well in experiments, will be readily deployable to the actual industrial control systems.

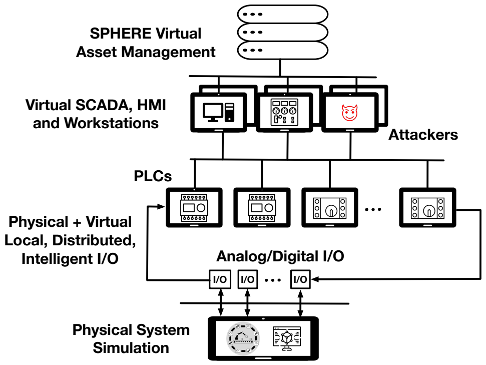


Figure 2: ICS study enabled by SPHERE.

Studying IoT behavior and its privacy implications. SPHERE will enable researchers to study private information leakage from black-box IoT devices (e.g., Alexa, Google Home), running proprietary software.

Researchers will be able to configure these devices and interact with them remotely, then analyze network traffic captured by SPHERE to identify any private information being sent to the device manufacturer's cloud. Because SPHERE will host different IoT devices in the same category (e.g., smart speakers), it will enable comparison between devices and between their communication protocols. SPHERE will also enable researchers to study cross-device information leakage, when IoT devices are allowed to exchange information.

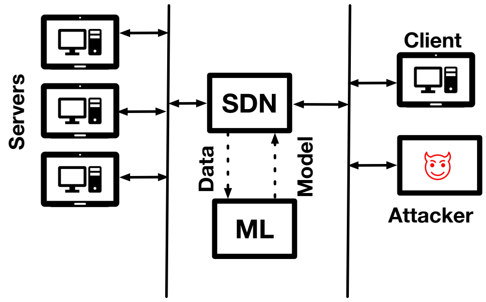


Figure 3: AI-enhanced attack mitigation enabled by SPHERE.

Studying AI-enhanced network attack detection and mitigation. Today's CS&P research increasingly focuses on attack-agnostic detection and mitigation (e.g., [46, 54, 55, 56]). SPHERE will facilitate exploration of the real-time deployability and scalability of machine learning in networks and lead to realistic evaluation of the proposed network attack defenses. It will enable experiments with general purpose compute nodes (for traffic generation), machine-learning nodes (for learning in real time) and programmable switches (for mitigation implementation). Researchers will be able to test dynamic scenarios in which network traffic flows through the programmable switch, where it is captured and redirected to machine-learning nodes for training or classification. These nodes then update the switch with dynamic mitigation rules. We illustrate a SPHERE-enabled experimentation scenario in Figure 3.

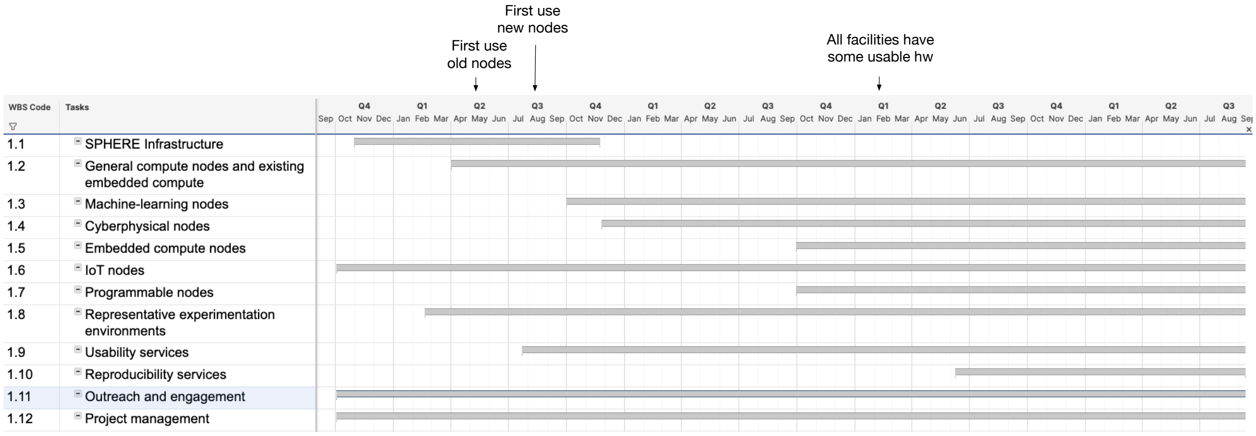
Studying embedded system security. SPHERE will provide heterogeneous edge white-box compute devices, including those found in tablets and mobile phones. It will allow researchers to create a single experiment, involving hundreds of edge devices, and involving some general compute or machine learning nodes to simulate cloudlets. Such experiments can be used to study security and scalability of distributed consensus and federated learning protocols, while replicating representative device join and leave scenarios, and representative network conditions, and while honoring unique resource constraints of edge devices.

Evaluation at different levels of fidelity. SPHERE blends emulation and experimentation with real, specialized hardware to accommodate a range of different research use cases. For example, general purpose nodes will commonly run experiments where entities such as servers, switches, routers, and edge devices are emulated on general compute nodes, and interconnected using virtual overlay networks. However, SPHERE will also incorporate specialized hardware in several different domains (e.g., programmable switches, embedded compute and IoT nodes, PLCs) that can be used for research that is sensitive to precise device-level features, access modalities, or performance characteristics. This flexibility will allow SPHERE to address the needs of different researchers, but also to address the needs of a single given project as it matures over time from emulation to real device deployment (e.g., from Mininet to real SDN switch).

5.7 Preliminary Activities Accomplished

As noted in Section 6, our team has built and operated the only public testbed for cybersecurity experimentation for almost two decades. We have developed the Merge control software [42] for experimental environments, deploying it over the four years to successfully run three different research infrastructures – DCOMP, Lighthouse and DeterLab. General compute nodes from DeterLab and embedded compute nodes from DCOMP will be absorbed by SPHERE. Thus we have the initial control software and seed hardware needed to start the buildout of SPHERE.

We ran the following efforts to identify scientific needs — Cybersecurity Experimentation of the Future (CEF) study groups in 2014, community engagement meetings in 2018, and workshop in 2022, and a Cybersecurity Artifacts Workshop in 2022. We have publicly released reports from each event [4, 5, 7]. We



WBS Code	Tasks
1.1	SPHERE Infrastructure
1.2	General compute nodes and existing embedded compute
1.3	Machine-learning nodes
1.4	Cyberphysical nodes
1.5	Embedded compute nodes
1.6	IoT nodes
1.7	Programmable nodes
1.8	Representative experimentation environments
1.9	Usability services
1.10	Reproducibility services
1.11	Outreach and engagement
1.12	Project management

Figure 4: Project schedule

have also surveyed publications from top cybersecurity and privacy venues in 2022 and produced a report of their experimental approaches [8]. These efforts informed our understanding of scientific needs.

In planning for SPHERE, we have developed an initial design of how experimental nodes and infrastructure would be distributed over the four facilities (Equinix DC, USC DC, USC-ISI DC and NEU IoT Lab) and racked, and we have verified that we can meet power and space restrictions of each facility. This design is provided as supplemental document.

5.8 Implementation Plan

Figure 4 shows our proposed project schedule. Deploying many different classes of nodes in a single infrastructure is a significant implementation challenge. We are confident that we can meet this challenge, because of our prior experience in running large research infrastructures, our reliance on the mature Merge software, and our careful project schedule. *Prior experience:* USC-ISI has already deployed and successfully run research infrastructures with general and embedded compute nodes using the Merge software (DCOMP and DeterLab). NEU has deployed and used for their private research various IoT devices in their locally-accessible Mon(IoT)r lab. We will leverage much of our RI design knowledge and the control-software implementation path (implementation scripts, architecture, node imaging and control approach) in our proposed work to deploy new general compute, embedded and IoT nodes. Machine-learning nodes are similar to general compute nodes, and should only require slight modifications to Merge implementation path. *Reliance on Merge:* Merge software is uniquely designed to support easy addition of new resource types to research infrastructure. *Project schedule:* We have spread the implementation tasks over four years, with hardware infrastructure being built out gradually in years 1, 2 and 3. Our schedule ensures that each node class is integrated and fully tested, before we proceed to add another class. CPS, programmable nodes and IoT nodes require more significant Merge code development, as well as careful architecture design, and IoT nodes require additional installation and management scripts for new node types, IoT sensors and IoT actuators. We have planned sufficient time in years 2 and 3 of the project to ensure integration of these nodes, with physical buildout of IoT infrastructure at NEU starting at year 1. We have also ensured that code-development and functionality-development tasks that are necessary for the usability of SPHERE will be developed early, so that we can start interacting with initial users, and incorporate their feedback as the project continues. Code and functionality that enhance SPHERE, but are not critical for usability, are pushed into year 4.

SPHERE will support the notion of multi-site experiments, where nodes from multiple specialized hardware platforms are composed into an integrated experiment. While this is challenging in general, our approach builds on the existing mature Merge testbed platform, which already provisions resources and network substrate at one datacenter (site) using VXLAN overlays that segment user traffic at Layer-2 and provides isolation between users. In SPHERE, we will build on Merge to extend these overlays across datacenters. Merge already leverages industry standard control plane algorithms such as MP-BGP EVPN (multiprotocol BGP Ethernet VPN), which are commonly adopted in geo-distributed industry cloud platforms to stretch VXLAN overlays across arbitrary routed substrates. Such substrates can include public Internet in the general case, but dedicated links between sites are preferred, providing better and more consistent performance than is possible over public links. We are thus confident that we can support integrated experiments that draw resources from any datacenter in SPHERE.

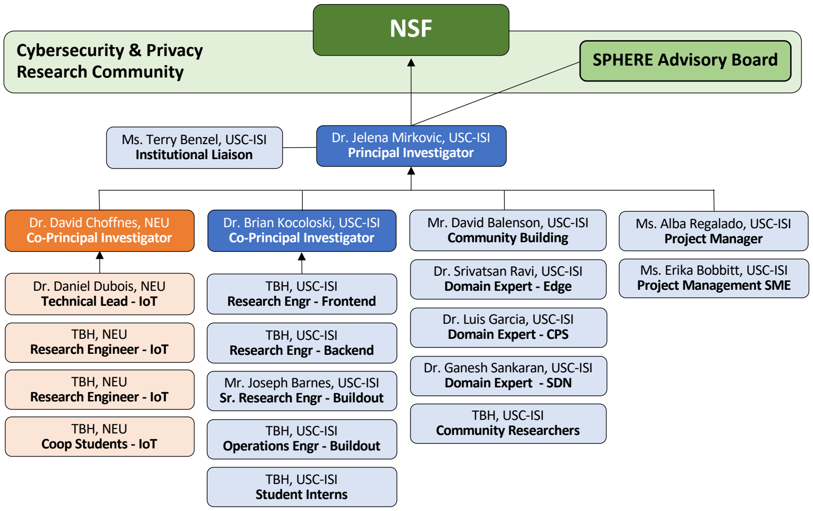

NSF
Cybersecurity & Privacy Research Community
SPHERE Advisory Board
Ms. Terry Benzel, USC-ISI Institutional Liaison
Dr. Jelena Mirkovic, USC-ISI Principal Investigator
Dr. David Choffnes, NEU Co-Principal Investigator
Dr. Brian Kocoloski, USC-ISI Co-Principal Investigator
Mr. David Balenson, USC-ISI Community Building
Ms. Alba Regalado, USC-ISI Project Manager
Dr. Daniel Dubois, NEU Technical Lead - IoT
TBH, USC-ISI Research Engr - Frontend
Dr. Srivatsan Ravi, USC-ISI Domain Expert - Edge
Ms. Erika Bobbitt, USC-ISI Project Management SME
TBH, NEU Research Engineer - IoT
TBH, USC-ISI Research Engr - Backend
Dr. Luis Garcia, USC-ISI Domain Expert - CPS
TBH, NEU Research Engineer - IoT
Mr. Joseph Barnes, USC-ISI Sr. Research Engr - Buildout
Dr. Ganesh Sankaran, USC-ISI Domain Expert - SDN
TBH, NEU Coop Students - IoT
TBH, USC-ISI Operations Engr - Buildout
TBH, USC-ISI Community Researchers
TBH, USC-ISI Student Interns

Figure 5: Organizational chart.

SPHERE RI will first open to users in the third quarter of year 1, with general compute and embedded compute nodes absorbed from DeterLab and DCOMP. New nodes will be first usable in the fourth quarter of year 1. All node types will have at least some nodes available for use in the second quarter of year 3. Because this is an implementation project, we did not budget resources for operation and user support in general. However, SPHERE RI will be able to accommodate users starting in the third quarter of year 1. We will obtain SPHERE users in the following ways: (1) we will transition DCOMP and DeterLab users to SPHERE as we absorb nodes from these research infrastructures, (2) we will organize need-discovery events at top cybersecurity and privacy conferences, and will invite researchers to try SPHERE, (3) we will work with artifact contributors and evaluators to help them use SPHERE. We will work with SPHERE users to draft a small team of “beta users”. We will work closely with beta users to gather feedback on our implementation, and input for future tasks. Our schedule includes regular software and hardware maintenance, which is necessary to keep the RI in operational state. Our schedule also includes user support for beta testers, for contributors of REEs from the researcher community, and for research artifact evaluation committees.

Our organizational chart is shown in Figure 5, and key personnel and their roles are summarized in Table 6, with more details available in PEP.

Prior to the full project award, we will create an advisory board composed of several government, industry, and academic senior leaders and visionaries from diverse communities that are relevant to CS&P research infrastructure and experimentation. We will seek advice and support from the advisory board throughout the life of the project. Our success metrics for implementation tasks include matching the progress of the project with the integrated master schedule, achieving operational status in year 1, and attracting annually at least 10 new research users for each node class, and at least 100 new research users for the entire RI.

Person	Role	Brief Description
Dr. Jelena Mirkovic	Principal Investigator	Lead for overall project direction, design, development, integration, and test
Ms. Terry Benzel	Institutional Liaison	Division Director and cross institute and university coordination
Dr. Brian Kocoloski	Co-Principal Investigator	Technical lead for overall project design, development, integration, and test
TBH1	Research Engineer	Backend development
TBH2	Research Engineer	Frontend and backend development
Mr. Joseph Barnes	Senior Research Engineer	Lead for equipment and buildout
TBH3	Operations Engineer	Buildout
Mr. David Balenson	Community Building	Community outreach and engagement
Dr. Srivatsan Ravi	Domain Expert	Domain expert and liaison to researchers for edge computing
Dr. Luis Garcia	Domain Expert	Domain expert and liaison to researchers for CPS
Dr. Ganesh Sankaran	Domain Expert	Domain expert and liaison to researchers for SDN
Ms. Alba Regalado	Project Manager	Project admin, planning, monitoring, reporting
Ms. Erika Bobbitt	Project Management SME	Project management guidance and advice
Dr. David Choffnes	Co-Principal Investigator	Lead for IoT design, development, integration, and test at NEU
Dr. Daniel Dubois	Research Scientist	Technical lead for IoT development at NEU
TBH4	Research Engineer	Developer for IoT network, data, and integration
TBH5	Research Engineer	IoT buildout lead and developer for control

Table 6: Key personnel and their roles on the project.

5.9 Operations and Utilization Plan

Management plan: SPHERE will be managed by the PIs and our project management team, and available to all U.S. researchers and educators. The infrastructure will be accompanied by continued user outreach and engagement activities, and will include training and opportunities for ongoing contributions from students and underserved populations. We have many years of experience in operating research infrastructure and obtaining funding through several government agencies. We expect that operations funding will come from a mix of NSF programs in OAC and CISE and possible research communities, such as DARPA and the DOE.

Governance, user access and utilization approach: Access to SPHERE will be governed by a set of standard rules of engagement. Users will access SPHERE via a single user interface, and authenticate through a federated login system or via SPHERE-specific credentials. Users will interact with SPHERE by choosing one of its six user portals, and switching if their needs evolve. Users will request resources on demand, leasing them for a limited time period. A lease can be extended, as long as no other users request the same resource. We will also offer an option to reserve resources for future use. Any resource contention will be resolved by temporarily stopping lease extensions for the given resource type.

Success metrics: As part of the implementation effort we propose a set of success metrics for the operational life-cycle, which may be refined when the operation starts. The following metrics would be reported separately for research users and for education users to measure *adoption, utilization and user diversity*: (1) number of users (total, new and active), (2) user demographics, (3) user institutions per category (U.S. vs international, MSI vs non-MSI, 2-year vs 4-year, tier-1 vs other), (4) utilization per node class and overall. The following metrics would be reported for the entire RI to measure *research impact*: (1) research publications using SPHERE, (2) research artifacts on SPHERE (contributed by AECs and contributed by users), (3) artifact reuse attempts, (4) REE contributions, (5) REE use attempts. The following metrics would be reported to measure *RI alignment with user needs*: (1) user satisfaction, measured through periodic surveys integrated with our user portals, (2) help tickets, (3) ticket response time.

6 Team Qualifications

Our team has extensive prior work in supporting and promoting representative, reproducible cybersecurity experimentation. The USC-ISI team has designed, implemented and operated DeterLab [27, 28] –

currently the only public research infrastructure for cybersecurity – for over 19 years (funded by numerous NSF, DARPA and DHS grants). This infrastructure has served a broad research community, including 389 research projects, from 278 institutions and involving 1,042 researchers, from 205 locations and 46 countries. In addition to hardware infrastructure, our team has developed many innovations to make the experimentation process more scientific, representative and reproducible. These include realistic traffic generation tools [57, 58, 59], realistic network emulators [60], human user behavior models [61], benchmarks for DDoS defenses [62], experiment automation tools, such as MAGI [63] and DEW [38], experiment logging and monitoring tools [52, 64], flexible experiment traffic containment tools [50], etc. The USC-ISI team developed the Merge control software [42] for experimental environments, deploying it over the four years to successfully run three different research infrastructures – DCOMP, Lighthouse and DeterLab.

We have further pioneered use of public testbeds in CS&P education. PI Mirkovic led two NSF-funded projects (award number #0920719 and #1224035), which developed materials for online, hands-on cybersecurity teaching using DeterLab testbed. These materials have so far been used by 237 classes, from 149 institutions and impacting 20,365 students.

To support reproducible research, we have developed an online catalogue for cybersecurity artifacts, called SEARCCH [65] (NSF award #1925616). For more than 15 years, we have been leaders on the the steering committee for Cyber Security Experimentation and Test (CSET) workshop. We led Cybersecurity Experimentation of the Future (CEF) study groups in 2014, community engagement meetings in 2018, and a workshop in 2022, and a Cybersecurity Artifacts Workshop in 2022. Terry Benzel, Senior Personnel and Institutional Liason for SPHERE, brings up-to-date insights from her participation on the NSF CISE Advisory Board, the FABRIC Advisory Board and the IEEE Computer Society Board of Governors.

Northeastern University has built the largest private IoT testbed in the world, the Mon(IoT)r Lab [66] in 2017, with 123 IoT devices currently deployed. The Mon(IoT)r testbed is a first-of-its-kind “living lab”, set up like a studio apartment where students, faculty, and staff can run controlled experiments and observe IoT device behavior while 98 volunteer subjects enrolled in an IRB-approved study use the IoT devices as they see fit. Software and documentation for building our Mon(IoT)r testbed have been published [67] and used by other institutions, including Imperial College London, University College London, Politecnico di Torino, and University of California Irvine. Our IoT Lab efforts have led to several publications in top conferences, including IMC [68, 69, 70], PETS [71, 72], ACNS [73], S&P [74] and CHI [75]. The data produced by our IoT devices has been publicly shared and downloaded 559 times. Our research results have also been presented in the FTC PrivacyCon 2020 and 2022, and featured by the popular press, including The New York Times [76, 77], The Financial Times [78], USA Today [79], and Consumer Reports [80, 81, 82].

7 Institutional Commitment to Diversity and Inclusion

Both *USC* and *Northeastern University* have a dedicated *Office of Diversity, Equity and Inclusion* [83, 84] to promote the university’s commitment to equal opportunity, affirmative action, diversity, and social justice, while building a climate of inclusion. Additionally, *USC’s Viterbi School of Engineering*, engages in a wide range of initiatives to promote diversity and inclusion, including the Center for Engineering Diversity, the K-12 STEM Center, the Viterbi Summer Institute, JEDI Peer Educator program, and USC Gateway program. The *USC CS department* has its own DEI initiatives, specifically focused on increasing representation of students applying to it, and increasing the retention rate of Hispanic/Latino and African-American students. *USC-ISI* has its own DEI team, engaging directly with USC-ISI community. USC-ISI has launched a new Research Engineer Internship Program in 2023, to develop a pipeline of talent from masters and undergraduate students. USC-ISI has also started Stimulating STEM Summer program in 2022, a free four-week residential summer program, to stimulate interest in STEM among students from marginalized communities in Los Angeles. *PI Mirkovic* is the PI for the NSF-funded REU site at ISI (NSF

Awards #1659886 and #2051101), which has run for five years so far, and has involved 50 undergraduate students (25 from under-resourced institutions) in computer science research. *Northeastern University* DEI initiatives include learning programs (e.g., the Align program [85], a direct pathway to a MS in CS for all undergraduate backgrounds), funding opportunities (e.g., Khoury Research Apprenticeship [86] to promote multidisciplinary student research), students clubs, and engagement with K-12 pipeline (e.g., NEU partnership with SMASH Academy [87]). *Khoury College of Computer Sciences at NEU* has its own DEI initiatives, aimed to increase participation of women and BIPOC students at all levels of CS education.

8 Divestment

SPHERE physical systems and the software infrastructure that enable the community to use the research infrastructure make up the full SPHERE system. It is expected that SPHERE will be transitioned to sustainable operations as described in SPHERE Operational Plans and Projected Cost, Supplementary Document. Ultimately divestment and disposition will be the responsibility of the commissioning and operations team for SPHERE. We do not anticipate the need for any disposition during the course of this Mid-Scale Research Infrastructure-1 project. All computational systems purchased on the project will vest with and ultimately be disposed of by the acquiring organizations (University of Southern California and Northeastern University). The universities have established equipment disposition policies and practices. All software developed during the SPHERE project will be available in open-source repositories. We do not anticipate any unusual conditions that would impact disposition.

9 Results from Prior NSF Support

PI Mirkovic is lead PI on NSF award #2016643 “ENS: Modernizing and Streamlining DeterLab Testbed Experimentation” ($2,000,000, 8/2020 - 8/2023). *Intellectual Merit.* This project modernizes the community RI of the public DeterLab testbed, by adding new, modern nodes and switches, which support high-density virtualization, and by replacing old Emulab RI control software with Merge software. The modernization was completed at the start of 2023. The effort further streamlines experimentation by offering new tools for CS&P experiments, such as new traffic generators and human user simulators. No publications were yet produced under this award; data products are at [88, 89, 90, 91, 92]. *Broader Impacts.* DeterLab’s modernization has transformed experimentation for cybersecurity and privacy communities, offering abundant compute resources, and unprecedented speed and reliability of resource commissioning. DeterLab is currently in use by thousands of researchers and students, the majority of which come from underserved communities. All DeterLab users benefit from easier resource allocation and more streamlined experimentation. *Relationship to proposed work.* The new portions of DeterLab testbed will be absorbed by the proposed SPHERE RI.

Co-PI Choffnes is a PI on NSF award #1955227 “SaTC: Frontiers: Collaborative: Protecting Personal Data Flow on the Internet” ($1,700,000, 10/2020 - 9/2025). *Intellectual Merit.* This project, ProperData, entails a multidisciplinary, multi-modal approach to understanding privacy implications of online activities, designing systems to improve this privacy, and engaging with experts from law and economists to effect change in policies regarding personal data protection. *Broader Impacts.* The ProperData project has led to more than a dozen publications in peer-reviewed conferences and data products, all listed at [93]. It has produced engagement with public policy via the annual Federal Trade Commission PrivacyCon event and a collaboration with Consumer Reports, resulted in numerous articles in the popular press, and hosted two workshops on privacy (contact tracing apps and IoT devices). *Relationship to proposed work.* While the Frontiers project supports research on IoT privacy, it does not build infrastructure for others to run experiments in CS&P as envisioned by the proposed work.

Co-PI Kocoloski has no prior NSF funding.

References Cited

[1] NSF. NSF's 10 Big Ideas. https://www.nsf.gov/news/special_reports/big_ideas/.

[2] Young, S. D. & Lander, E. S. Multi-Agency Research and Development Priorities for the FY 2023 Budget (2022).

[3] Cybersecurity and Information Assurance Interagency Working Group, National Science and Technology Council. FY 2023 Federal Cybersecurity R&D Strategic Plan Implementation Roadmap. https://www.nitrd.gov/pubs/FY2023-Cybersecurity-RD-Roadmap.pdf.

[4] Balenson, D., Tinnel, L. & Benzel, T. Cybersecurity experimentation of the future (CEF): catalyzing a new generation of experimental cybersecurity research. *SRI International, Tech. Rep.* (2015).

[5] Mirkovic, J., Balenson, D., Ravi, S., Garcia, L. & Benzel, T. Cybersecurity Experimentation Workshop – 2022 – Report. https://bit.ly/CyberExperWkshp2022 (2022).

[6] National Academies of Sciences, Engineering, and Medicine and others. Reproducibility and replicability in science (2019).

[7] Balenson, D. *et al.* Cybersecurity artifacts workshop – report. https://bit.ly/CyberArtifactsWkshp2022 (2022).

[8] Mirkovic, J. Survey of Experimentation Approaches in Cybersecurity and Privacy Papers. https://bit.ly/CyberPapersSurvey2022 (2022).

[9] Presidential Policy Directive – Critical Infrastructure Security and Resilience (PPD-21). https://obamawhitehouse.archives.gov/the-press-office/2013/02/12/presidential-policy-directive-critical-infrastructure-security-and-resil (2013).

[10] NPR. A 'Worst Nightmare' Cyberattack: The Untold Story Of The SolarWinds Hack. https://www.npr.org/2021/04/16/985439655/a-worst-nightmare-cyberattack-the-untold-story-of-the-solarwinds-hack.

[11] TechTarget.com. Colonial Pipeline hack explained: Everything you need to know. https://www.techtarget.com/whatis/feature/Colonial-Pipeline-hack-explained-Everything-you-need-to-know.

[12] Statista. Annual number of ransomware attacks worldwide from 2016 to first half 2022. https://www.statista.com/statistics/494947/ransomware-attacks-per-year-worldwide/.

[13] Government Technology. Hacktivism and DDOS Attacks Rise Dramatically in 2022. https://www.govtech.com/blogs/lohrmann-on-cybersecurity/hacktivism-and-ddos-attacks-rise-dramatically-in-2022.

[14] Sumeet Wadhwani, Spiceworks. Data Breaches Soared by 70% In Q3 2022 in an Otherwise Dull Year. https://www.spiceworks.com/it-security/data-security/news/data-breach-report/ (2022).

[15] Dark Reading. FBI, CISA: Cyber Actors Targeting COVID-19-Related Research. https://www.darkreading.com/attacks-breaches/fbi-cisa-cyber-actors-targeting-covid-19-related-research (2020).

[16] Dark Reading. Hypothesis: Cyber Attackers Are After Your Scientific Research. https://www.darkreading.com/edge-articles/hypothesis-cyber-attackers-are-after-your-scientific-research (2020).

[17] Ellen Nakashima and Devlin Barrett. U.S. accuses China of sponsoring criminal hackers targeting coronavirus vaccine research. Washington Post, https://www.washingtonpost.com/national-security/us-china-covid-19-vaccine-research/2020/07/21/8b6ca0c0-cb58-11ea-91f1-28aca4d833a0_story.html (2020).

[18] Kartikay Mehrotra. How Hackers Bled 118 Bitcoins Out of Covid Researchers in U.S. Bloomberg News, https://www.bloomberg.com/news/features/2020-08-19/ucsf-hack-shows-evolving-risks-of-ransomware-in-the-covid-era.

[19] Ravie Lakshmanan. Iranian Hackers Target High-Value Targets in Nuclear Security and Genomic Research. The Hacker News, https://thehackernews.com/2022/09/iranian-hackers-target-high-value.html.

[20] ALMA Observatory. ALMA Services Affected by Cyberattack. https://almaobservatory.org/en/announcements/alma-services-affected-by-cyberattack/ (2022).

[21] ALMA Observatory. ALMA Update on the Recovery from Cyberattack. https://almaobservatory.org/en/announcements/alma-update-on-the-recovery-from-cyberattack/ (2022).

[22] ALMA Observatory. ALMA Successfully Restarted Observations. https://almaobservatory.org/en/announcements/alma-successfully-restarted-observations/ (2022).

[23] Dark Reading. Cyberattackers Focus In on State-of-the-Art ALMA Observatory. https://www.darkreading.com/attacks-breaches/cyberattackers-focus-alma-observatory (2022).

[24] James Pearson and Christopher Bing. Exclusive: Russian hackers targeted U.S. nuclear scientists. Reuters, https://www.reuters.com/world/europe/russian-hackers-targeted-us-nuclear-scientists-2023-01-06/.

[25] Collberg, C. & Proebsting, T. A. Repeatability in computer systems research. *Commun. ACM* **59**, 62â69 (2016). URL https://doi.org/10.1145/2812803.

[26] Krafczyk, M. S., Shi, A., Bhaskar, A., Marinov, D. & Stodden, V. Learning from reproducing computational results: introducing three principles and the reproduction package. *Philosophical Transactions of the Royal Society* **379**, 20200069 (2021).

[27] Mirkovic, J. & Benzel, T. Deterlab testbed for cybersecurity research and education. *Journal of Computing Sciences in Colleges* **28**, 163–163 (2013).

[28] Wroclawski, J. *et al.* DETERLab and the DETER Project. In *The GENI Book*, 35–62 (Springer, 2016).

[29] Keahey, K. *et al.* Chameleon: a scalable production testbed for computer science research. In *Contemporary High Performance Computing*, 123–148 (CRC Press, 2019).

[30] Duplyakin, D. *et al.* The design and operation of {CloudLab}. In *2019 USENIX annual technical conference (USENIX ATC 19)*, 1–14 (2019).

[31] Breen, J. *et al.* POWDER: Platform for open wireless data-driven experimental research. In *Proceedings of the 14th International Workshop on Wireless Network Testbeds, Experimental evaluation & Characterization*, 17–24 (2020).

[32] Panicker, A. *et al.* Aerpaw emulation overview and preliminary performance evaluation. *Computer Networks* **194**, 108083 (2021).

[33] Raychaudhuri, D. *et al.* Challenge: COSMOS: A city-scale programmable testbed for experimentation with advanced wireless. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 1–13 (2020).

[34] Baldin, I. *et al.* FABRIC: A national-scale programmable experimental network infrastructure. *IEEE Internet Computing* **23**, 38–47 (2019).

[35] SAGE Cyberinfrastructure for AI at the Edge. SAGE is building a new kind of NSF Mid-scale Research Infrastructure (MSRI) that supports AI at the Edge. https://sagecontinuum.org/about/ (2019).

[36] Chameleon Team. Chameleon Documentation – Jupyter Interface. https://chameleoncloud.readthedocs.io/en/latest/technical/jupyter.html.

[37] FABRIC Team. Jupyter Examples. https://github.com/fabric-testbed/jupyter-examples/issues.

[38] DeterLab Team. DEW Portal. https://dew.isi.edu (2021).

[39] The DETER Project. Public Access to Shared Materials. https://www.isi.deterlab.net/sharedpublic.php.

[40] Chameleon Team. Trovi: Practical Open Reproducibility. https://chameleoncloud.gitbook.io/trovi/ (2022).

[41] CloudLab Team. CloudLab Documentation – Profiles. https://docs.cloudlab.us/basic-concepts.html.

[42] Goodfellow, R., Schwab, S., Kline, E., Thurlow, L. & Lawler, G. The DComp testbed. In *12th USENIX Workshop on Cyber Security Experimentation and Test (CSET 19)* (USENIX Association, Santa Clara, CA, 2019). URL https://www.usenix.org/conference/cset19/presentation/goodfellow.

[43] Producing an Ignition Config. https://docs.fedoraproject.org/en-US/fedora-coreos/producing-ign/ (2022).

[44] Zero Touch Provisioning - ZTP. https://docs.nvidia.com/networking-ethernet-software/cumulus-linux-44/Installation-Management/Zero-Touch-Provisioning-ZTP/ (2022).

[45] Ansible Documentation. https://docs.ansible.com/ (2022).

[46] Mirsky, Y., Doitshman, T., Elovici, Y. & Shabtai, A. Kitsune: an ensemble of autoencoders for online network intrusion detection. In *Proceedings of NDSS* (2018).

[47] Xu, Z., Ramanathan, S. S., Rush, A. M., Mirkovic, J. & Yu, M. Xatu: Boosting Existing DDoS Detection Systems Using Auxiliary Signals. In *Proceedings of the 18th International Conference on emerging Networking EXperiments and Technologies (CoNEXT 2022)* (2022).

[48] McGregor, I. Introduction to emulate3d: emulation, simulation, and demonstration. In *Proceedings of the Winter Simulation Conference*, 1–10 (Citeseer, 2012).

[49] Mattila, J., Ala-Laurinaho, R., Autiosalo, J., Salminen, P. & Tammi, K. Using Digital Twin Documents to Control a Smart Factory: Simulation Approach with ROS, Gazebo, and Twinbase. *Machines* **10**, 225 (2022).

[50] A. Alwabel and H. Shi and G. Bartlett and J. Mirkovic. Safe and Automated Live Malware Experimentation on Public Testbeds. In *Proceedings of CSET* (2014).

[51] Mirkovic, J., Bartlett, G. & Blythe, J. DEW: Distributed Experiment Workflows. In *11th USENIX Workshop on Cyber Security Experimentation and Test (CSET 18)* (2018).

[52] Mirkovic, J. *et al.* Using terminal histories to monitor student progress on hands-on exercises. In *Proceedings of the 51st ACM technical symposium on Computer Science Education (SIGCSE)*, 866–872 (2020).

[53] Hashi Conf. Automate image builds with Packer. https://www.packer.io/ (2021).

[54] Psathas, A. P., Iliadis, L., Papaleonidas, A. & Bountas, D. A hybrid deep learning ensemble for cyber intrusion detection. In *Proceedings of the 22nd Engineering Applications of Neural Networks Conference: EANN 2021*, 27–41 (Springer, 2021).

[55] Chalapathy, R. & Chawla, S. Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407* (2019).

[56] Zhao, H. *et al.* Multivariate time-series anomaly detection via graph attention network. In *2020 IEEE International Conference on Data Mining (ICDM)*, 841–850 (IEEE, 2020).

[57] Genevieve Bartlett and Jelena Mirkovic. Expressing Different Traffic Models Using the LegoTG Framework. In *Workshop on Computer and Networking Experimental Research using Testbeds (CNERT)* (2015).

[58] DeAngelis, D., Hussain, A., Kocoloski, B., Ardi, C. & Schwab, S. Generating representative video teleconferencing traffic. In *Proceedings of the 15th Workshop on Cyber Security Experimentation and Test*, 100–104 (2022).

[59] DeterLab Team. Flooder Case Study. https://docs.deterlab.net/orchestrator/flooder/l.

[60] Kline, E., Bartlett, G., Lawler, G., Story, R. & Elkins, M. Capturing Domain Knowledge Through Extensible Components. In Gao, H., Yin, Y., Yang, X. & Miao, H. (eds.) *Testbeds and Research Infrastructures for the Development of Networks and Communities*, 141–156 (Springer International Publishing, Cham, 2019).

[61] Blythe, J. & Tregubov, A. Farm: Architecture for distributed agent-based social simulations. In *International Workshop on Massively Multiagent Systems*, 96–107 (Springer, 2018).

[62] Mirkovic, J. *et al.* DDoS benchmarks and experimenter's workbench for the DETER testbed. In *2007 3rd International Conference on Testbeds and Research Infrastructure for the Development of Networks and Communities*, 1–7 (IEEE, 2007).

[63] Hussain, A., Jaipuria, P., Lawler, G., Schwab, S. & Benzel, T. Toward orchestration of complex networking experiments. In *CSET@ USENIX Security Symposium* (2020).

[64] Viswanathan, A., Hussain, A., Mirkovic, J., Schwab, S. & Wroclawski, J. A semantic framework for data analysis in networked systems. In *Proc. USENIX Symp. Netw. Syst. Des. Implement.(NSDI)*, 127–140 (2011).

[65] Balenson, D. *et al.* Toward Findable, Accessible, Interoperable, and Reusable Cybersecurity Artifacts. In *Cyber Security Experimentation and Test (CSET)* (2022).

[66] Choffnes, D. & Dubois, D. J. The Mon(IoT)r Lab. https://moniotrlab.khoury.northeastern.edu/ (2023).

[67] Dubois, D. J. & Choffnes, D. The Mon(IoT)r Lab Testbed Tools. https://moniotrlab.khoury.northeastern.edu/tools/ (2023).

[68] Ren, J. *et al.* Information Exposure for Consumer IoT Devices: A Multidimensional, Network-Informed Measurement Approach. In *Proc. of the Internet Measurement Conference (IMC)* (2019).

[69] Saidi, S. J. *et al.* A Haystack Full of Needles: Scalable Detection of IoT Devices in the Wild. In *Proc. of the Internet Measurement Conference (IMC)* (2020).

[70] Paracha, M. T., Dubois, D. J., Vallina-Rodriguez, N. & Choffnes, D. IoTLS: Understanding TLS Usage in Consumer IoT Devices. In *Proc. of the Internet Measurement Conference* (2021).

[71] Dubois, D. J. *et al.* When Speakers Are All Ears: Characterizing Misactivations of IoT Smart Speakers. In *Proc. of the Privacy Enhancing Technologies Symposium (PETS)* (2020).

[72] Mandalari, A. M. *et al.* Blocking without Breaking: Identification and Mitigation of Non-Essential IoT Traffic. In *Proc. of the Privacy Enhancing Technologies Symposium (PETS)* (2021).

[73] Shafqat, N. *et al.* ZLeaks: Passive Inference Attacks on Zigbee based Smart Homes. In *Proc. of the International Conference on Applied Cryptography and Network Security* (2022).

[74] Mandalari, A. M., Haddadi, H., Dubois, D. J. & Choffnes, D. Protected or Porous: A Comparative Analysis of Threat Detection Capability of IoT Safeguards. In *Proc. of the 44th IEEE Symposium on Security and Privacy (Oakland 2023)* (2023).

[75] Kowalcyzk, M. *et al.* Understanding Dark Patterns in Home IoT Devices. In *Proceedings of ACM Human Factors in Computing Systems (CHI 2023)* (Hamburg, Germany, 2023).

[76] Hill, K. Activate This ’Bracelet of Silence,’ and Alexa Can’T Eavesdrop. *The New York Times* (2020).

[77] NYT Editorial Board. Privacy Cannot Be a Casualty of the Coronavirus. *The New York Times* (2020).

[78] Murgia, M. Smart TVs sending private data to Netflix and Facebook. *Financial Times* (2019).

[79] Jolly, J. It’s not you, it’s them: Google, Alexa and Siri may answer even if you haven’t called. *USA Today* (2020).

[80] Waddell, K. Your Smart Devices Are Trying to Manipulate You With ‘Dark Patterns’. *Consumer Reports* (2023).

[81] Waddel, K. Connected Devices Share More Data Than Needed, Study Says. *Consumer Reports* (2021).

[82] John, A. S. Yes, Your Smart Speaker Is Listening When It Shouldn't. *Consumer Reports* (2020).

[83] USC. USC: Diversity, Equity and Inclusion. https://diversity.usc.edu/.

[84] Northeastern University. Office of Diversity, Equity and Inclusion. https://northeastern.edu/diversity/ (2023).

[85] Northeastern University. Align MS in Computer Science. https://www.khoury.northeastern.edu/programs/align-masters-of-science-in-computer-science/ (2023).

[86] Northeastern University. Khoury Research Apprenticeship. https://www.khoury.northeastern.edu/information-for-overview/current-masters-and-certificate-students/khoury-research-apprenticeship/ (2023).

[87] Kapor Center. SMASH Academy. https://www.smash.org/ (2023).

[88] DeterLab Team. DeterLab documentation. https://docs.deterlab.net/.

[89] Merge Team. Merge documentation. https://mergetb.org/docs/experimentation/.

[90] DeterLab Team. Modernized Deter Testbed. https://launch.mod.deterlab.net.

[91] Arasteh, S., Mirkovic, J. & Hauser, C. An Introduction to Dwarf. https://www.isi.deterlab.net/file.php?file=/share/shared/AnintroductiontoDwarf.

[92] Wei-Cheng Wu and Christophe Hauser. Binary Visualization Project. https://binary.deterlab.net.

[93] ProperData Team. ProperData: Protecting Personal Data Flow on the Internet. https://properdata.eng.uci.edu/publications/.