Terry Benzel
USC Information Sciences Institute
May 18, 2015

The Science of Cyber Security Experimentation
The DETER Project

- A research program:
  - To advance capabilities for experimental cybersecurity research

- A testbed facility:
  - To serve as a publicly available national resource…

- A community building activity:
  - To foster and support collaborative science
The DETER Facility

A general purpose, flexible platform for modeling, emulation, and controlled study of large, complex networked systems

Elements located at USC/ISI (Los Angeles), UC Berkeley, and USC/ISI (Arlington, VA)

Funded by NSF and DHS, started in 2003

Based on Emulab software, with focus on security experimentation

Shared resource – multiple simultaneous experiments subject to resource constraints

Open to academic, industrial, govt researchers essentially worldwide – very lightweight approval process
Physical Platform

- ~440 PC-based nodes
  - Berkeley, CA - ~200 Nodes
  - Los Angeles, CA - 220 Nodes
  - Arlington, VA - 20 Nodes

- Interconnect
  - 1 Gb/s - LA-UCB
  - 1-10 Gb/s LA-Arlington

- Local and Remote access
Research Goals

Advance our understanding of experimental cybersecurity science and methodologies
Enable new levels of rigor and repeatability
Transform low level results to high level understanding
Broaden the domains of applicability

Advance the technology of experimental infrastructure
Develop technologies with new levels of function, applicability, and scale

Share knowledge, results, and operational capability
Facility, data and tools
Community and knowledge
Scalable Modeling and Emulation

The problem:
Traditional testbeds can model and emulate small systems at a fixed level of fidelity.

The challenge:
Many real problems require modeling of large, complex systems at an appropriate (“good enough”) level of fidelity.
That level may be different for different parts of the modeled system.
Think of this as “smearing the computation power around to just where it’s needed”.
Containers

DETER containers use virtualization to support larger experiments.

Containers use several different types of virtualization.

Selecting different virtualization types allows a trade-off:
- One container per physical machine → high fidelity.
- More containers per physical machines → less fidelity.
Experiment scenarios require many disparate elements to be combined within a single overall scenario. These elements must be:

- deployed, initialized, configured,
- monitored and coordinated
- instrumented with real-time and post-mortem data collection

...throughout the execution of the experiment.

DETER’s MAGI agent infrastructure provides an architecture for scalable control and instrumentation
Multi-agent system to Model Some Human Behavior

Testbeds must model impact of human activity in repeatable experiments
   Provide more realistic behavior for testing security tools

   **But** real humans are expensive and non-repeatable

Model goal-directed team activity
   Measure impact of an attack on team goals

   Model impact of organization structure

Model certain human characteristics
   Propensity to make mistakes

   Aspects of physiology, (soon: emotion, bounded rationality)

   Flexibility to changing conditions

Configurable tool for experimenters
DETER User Institutions

**Government**
Air Force Research Laboratory
DARPA
Lawrence Berkeley National Lab
Naval Postgraduate School
Sandia National Laboratories

**Industry**
Agnik, LLC
Aerospace Corporation
Backbone Security

BAE Systems, Inc.
BBN
Bell Labs
Cs3 Inc.
Distributed Infinity Inc.
EADS Innovation Works

FreeBSD Foundation
iCAST
Institute for Information Industry

Intel Research Berkeley

**Academia**
Bar-Ilan University
Carnegie Mellon University
Columbia University
Cornell University
Dalhousie University
DePaul University
George Mason University
Georgia State University
Hokuriku Research Center
ICSI
IIT Delhi
IRTT
ISI
Johns Hopkins University
Lehigh University
MIT
New Jersey Institute of Technology
Norfolk State University
Pennsylvania State University
Purdue University
Rutgers University
Sao Paulo State University
Southern Illinois University
TU Berlin
TU Darmstadt
Texas A&M University
UC Berkeley

UC Davis
UC Irvine
UC Santa Cruz
UCLA
UCSD
UIUC
UNC Chapel Hill
UNC Charlotte
Universidad Michoacana de San Nicolas
Universita di Pisa
University of Advancing Technology
University of Illinois, Urbana-Champaign
University of Maryland
University of Massachusetts
University of Oregon
University of Southern California
University of Washington
University of Wisconsin - Madison
USC
UT Arlington
UT Austin
UT Dallas
Washington State University
Washington University in St. Louis
Western Michigan University
Xiangnan University
Youngstown State University
Education

Hands on exercises
Students gain from direct observation of attacks and interaction
Pre packaged for both student and teacher
  - Buffer overflows, command-injection, man-in-the-middle, worm modeling, botnets, and DoS

Facility support for class administration
Next Steps - Ecosystem

SRI and ISI developing strategic plan:
Planning grant from NSF

Study current/ future cybersecurity research

Current/expected experimentation infrastructure

Create roadmap for developing:

Accessible, broad, and multi-organizational cybersecurity experimentation capability that meets tomorrow’s research needs
The Road to Tomorrow: Cybersecurity Experimentation of the Future
Motivation: Why are We Doing This?

Society’s cyber dependencies are rapidly evolving.
In nearly every aspect of our lives, we are moving toward pervasive embedded computing with a fundamental shift in network properties.
These changes bring a very real and wide-ranging set of challenging cyber threats.
Addressing these challenges will require cybersecurity research based on sound scientific principles.
The scale and complexity of the challenges will require that researchers apply new experimentation methods that enable discovery, validation, and ongoing analysis.
Cybersecurity R&D is still a relatively young field
It involves intrinsically hard challenges
   Inherent focus on worst case behaviors and rare events
   In the context of multi-party and adversarial/competitive scenarios

Research infrastructure is crucial
   Allow new hypotheses to be tested, stressed, observed, reformulated, and ultimately proven before making their way into operational systems

Ever increasing cyber threat landscape demands new forms of R&D and new revolutionary approaches to experimentation and test
Clearly a need for future research infrastructure that can play a transformative role for future cybersecurity research
The Need for Transformational Progress

Transformational progress in three distinct, yet synergistic areas is required to achieve the desired objectives:

1) Fundamental and broad intellectual advance in the field of experimental methodologies and techniques
   With particular focus on complex systems and human-technical interactions

2) New approaches to rapid and effective sharing of data and knowledge and information synthesis
   That accelerate multi-discipline and cross-organizational knowledge generation and community building

3) Advanced experimental infrastructure capabilities and accessibility

A Science of Cybersecurity Experimentation
Science of Cybersecurity Experimentation

New direction for the field of experimental cybersecurity R&D
R&D must be grounded in scientific methods and tools to fully realize the impact of experimentation
Different than and complementary with the science of cybersecurity

- New approaches to sharing all aspects of the experimental science – data, designs, experiments, and research infrastructure
- Cultural and social shifts in the way researchers approach experimentation and experimental facilities
- New, advanced experimentation platforms that can evolve and are sustainable as the science and the community mature

Source: https://www.nsa.gov/research/tnw/tnw192/article4.shtml
Roadmap for a New Generation of Experimental Cybersecurity Research

The roadmap presents requirements, objectives and goals for 30 key capabilities organized into 8 core areas over 3, 5, and 10 year phases. Some phases build upon each other and others require new fundamental research over a long time period.

- Key capabilities consider:
  - Current experimental cybersecurity research and its supporting infrastructure
  - Other types of research facilities
  - Existing cyber-domain “T&E” capabilities (primarily DoD)

- The roadmap presumes advances in key computer science disciplines
A Definition of “Cybersecurity Experimentation Infrastructure”

General purpose ranges and testbeds (physical and/or virtual)
Specialized ranges and testbeds (physical and/or virtual)
Software tools that supports one or more parts of the experiment life cycle, including, but not limited to:
   Experiment design
   Testbed provisioning software
   Experiment control software
   Testbed validation
   Human and system activity emulators
   Instrumentation - systems and humans
   Data analysis
   Testbed health and situational awareness
   Experiment situational awareness
   Other similarly relevant tools

Specialized hardware tools - simulators, physical apparatus, etc.
The goal is not to create a single instance of a cyber experimentation testbed or facility. Over time the roadmap may be realized through an ecosystem of many different instantiations – from small, stand-alone and localized to large distributed experimental capabilities, all spanning multiple domains.
Hybrid Architectures Based on Different Building Blocks

Cloud technology
Software defined networking (SDN)
Knowledge sharing and community environments
  E.g., Eclipse

Emulated and simulated environments
  E.g., RTDS, wireless

Specialized hardware
  E.g., FPGA, GPU, Intel Xeon Phi

No single hardware/software substrate
Research Infrastructure is More than Infrastructure

Research infrastructure >> infrastructure of machines and tools
Scientific methodologies, experimental processes, and education are critical to effective use of the machines and tools.

Research infrastructure requires meta-research into:
- Design specification (multi-layered languages and visualization)
- Abstraction methodologies and techniques
- Semantic analysis and understanding of experimenter intent
- Formal methods and a rich approach to modeling to satisfy science objectives
Where is Experimentation Applicable?

Overarching goal is to increase researcher effectiveness and support the generation and preservation of solid empirical evidence

- Infrastructure to enable research, not constrain
- New mechanisms to capture and share knowledge (designs, data and results) to enable peer review and allow researchers to build upon each other

Experimentation is about learning
- To perform an evaluation (not formal T&E)
- To explore a hypothesis
- To characterize complex behavior
- To complement a theory
- To understand a threat
- To probe and understand a technology
Representative Cybersecurity Hard Problems

Systems/software
  Human interactions

  System of system security metrics

  Emergent behavior in large scale systems

  Supply chain and root of trust

  Societal impacts and regulatory policies

Networking
  Anonymity and privacy of data and communication

  Trust infrastructure

  Software defined networking (SDN)

  Political, social, and economic (balance-of-interest) goals in network design

  Pervasive communications, across organizational and political boundaries

  Cyber-physical systems
    - Embedded devices
    - Autonomous vehicles, smart transportation
    - Electric power, smart grid
    - Medical implants, body sensors, etc.
Experimentation - It’s About the Real World

Experimentation should start with models of the real world. Modeling and abstraction allow us to capture conceptual models of the real world with varying degrees of fidelity. Key research areas include:

- Experiment design specifications
- Auto-generated model refinement
- Methodologies and tools to assess representation
  - Understanding the multiple dimensions of realism
  - Taxonomies of realism metrics for realism sufficiency
  - Additional methods and tools can help extend these modeling activities to provide increasing forms of real world models
Leveraging Existing Infrastructure

Leverage current NSF, DHS, and DOD investments as starting points.

**GENI**
- Community model
- Integration of real infrastructure (overlay network, part of real deployments)

**DETER**
- Research and operational knowledge
- Experimental framework
- Specialized experimentation and cybersecurity tools
  - E.g., MAGI, Containers

**MIT Lincoln Laboratory**
- Standard experiment specification languages
  - E.g., CRIS, CCER
- Specialized experimentation tools
  - E.g., ALIVE, LARIAT, LLAF, KOALA

Other government sponsored infrastructure – DOD, DOE, DOT, etc.
Join Us - students, post docs, research programmers, computer scientists, faculty