Science of Security Experimentation

John McHugh, Dalhousie University
Jennifer Bayuk, Jennifer L Bayuk LLC
Minaxi Gupta, Indiana University
Roy Maxion, Carnegie Mellon University
Moderator: Jelena Mirkovic, USC/ISI
Topics

• **Meaning of science**

• **Challenges to rigorous security experimentation:**
  – **Approach?** choice of an appropriate evaluation approach from theory, simulation, emulation, trace-based analysis, and deployment
  – **Data?** how/where to gather appropriate and realistic data to reproduce relevant security threats
  – **Fidelity?** how to faithfully reproduce data in an experimental setting
  – **Community?** how to promote reuse and sharing, and discourage reinvention in the community

• **Benchmarks?** Requirements for and obstacles to creation of widely accepted benchmarks for popular security areas

• **Scale?** When scale matters?
Top Problems

• Good problem definition and hypothesis
  – Lack of methodology/hypothesis in publications
  – Learn how to use the word “hypothesis”

• Lack of data
  – Data is moving target, hard to affix science to attacks that change

• Program committees
  – Hard to publish, hard to fund, no incentive to good science
  – Data needs to be released with publications

• Who really cares except us?

• Rigor applied to defenses not to attacks
  – Define security

• Do we want science or engineering?

• Years behind attackers

• Provenance, tools that automate collection of provenance
Closing statements

• Learn from publications in other fields
• What you did, why was it the best thing to do (methodology and hypothesis matter)
• Right now we have the opportunity to change
  – Learn from other fields before we grow too big too wide too fast
  – We must avoid adopting wrong but easy approaches, hard to change
• Data is crucial, we need to focus on getting more data on ongoing basis
  – One-off datasets don’t cut it
Approach

• Use what you think will give you the best answer for the question you have
  – Understanding your options and your hypothesis is what matters, the rest is given
  – Also constraints on time and resources

• Write up all the details in the methods section
  – Forcing people to write this all down would lead to many paper rejections and would quickly teach people about the rigor
  – Experience with QoP shows it’s hard to even have people write this down, let alone do it correctly
Data

- Who has the data?
- How to get access?
- Lengthy lawyer interactions. In the meantime research isn’t novel anymore.
- Resources to store data
- Results cannot be reproduced when data is not public
- No long-term data sets (10 years, study evolution) in real time
  - Need good compute power where the data is
  - There are common themes in data analysis – this could be precomputed
- [www.predict.org](http://www.predict.org) (lots of data here)
- Hard to get data on attacks before persecution is done, may be years. Also companies don’t want to admit to be victims.
Data

• Metadata necessary for usefulness (anonymization, limitations, collection process)
  – Not enough info to gauge if data is useful to researchers
  – No detail about sanity checks, calibration steps
  – Improve collection design AND disclose it
• Understanding of common data products would drive better collection rigor
• Not every question can be answered with a given data
  – relationship of data to problems is important
• Provenance on data, what can be done with it
• Keystroke data with proper metadata (by Roy Maxion)
  – http://www.cs.cmu.edu/~keystroke
Community

- We’re competing among each other, attackers are advancing
- Adoption of protocols is field for research
- Problems that lack datasets are just not being addressed
- Teaching builds better experimental practices
  - Requirement courses for degrees
- Rigor requirements in conflict with funding
  - Actually in conflict with publishing and research community
Meaning of Science

• Tightly focused question
  – Forming a research hypothesis
    • Then validity, reproducibility by someone else, repeatability - are important
    • Repeatability – same run similar answers
    • Validity
      – External validity - can you generalize your claims to a different, larger, population
      – Internal validity – logical consistency internally in the experiment

• There’s no building on work of others so rigor is not necessary
  – We don’t even have the right questions formed

• NSF workshop on science of security, Dec’08 in Claremont
Where to Start?

• Formulating good questions
  – Predictability is a hard problem in security
  – Well-defined, small, constrained problems make sense
• Take courses on experimental design/methodology (students)
• Read papers and critique the methodology in them
• Finding right tools to produce answers
Where to Start?

• Security means different things to different people
  – Must define which attribute of security you’re measuring

• What PC’s could do:
  – Enforce methodology/hypothesis questions
  – Enforce reproducibility

• Extra work with no quick payoff for select few that do what we suggest

• Attackers can avoid well-defined models
  – We need stronger models then
Where to Start?

• Attackers are evolving – moving target
  – Hard to match this pace with methodology evolution
  – Major logic is missing

• Large number of things manifest as security problems but are not
  – Buffer overflows are coding problems, sloppy sw
What to Fund

• Education
• A critical review journal
• Requirements analysis
  – Attributes of systems that give you assurance that your goals are met
  – Close specification of context