Is it Possible to Do Scientific Experiments in Security?

Roy A. Maxion

Dependable Systems Laboratory
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213
Email: maxion@cs.cmu.edu

Science of Security Workshop
National Science Foundation
18 November 2008
Berkeley, California

Science of security - I

- First, since we are talking about science, let's review some of its precepts.
- Science derives laws that explain principles operating in nature ... by ...
  - Observing (experimenting with) phenomena of interest
  - Controlling the risk of bias in those observations such that they are reliable, repeatable and valid
  - Predicting future observations on the basis of present ones (i.e., generalizing from the derived laws)
  - Eliminating alternative explanations
  - Explaining causal mechanisms

Science of security - II

- Science ...
- ... comprises knowledge covering general truths, i.e., the operation of general laws.
- ... deals with objectively measurable phenomena
- ... predicts ...by virtue of having laws ...
- ... generalizes, largely by asking questions about the conditions under which the laws apply.
- The discovery of those laws is usually done by experiment.

Charge to the speakers ...

- **What makes a good security experiment?**
- **What can and cannot be learned about security through experiments?**
- **Should there be better connections between formal and experimental security work?**
- **How can we improve the state-of-the-art for computer security experiments?**

(For want of time, I will address only the first and last of these.)

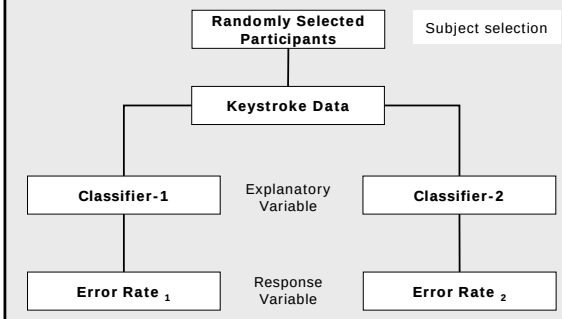
Charge to the speakers ...

- **What makes a good security experiment?**
- **What can and cannot be learned about security through experiments?**
- **Should there be better connections between formal and experimental security work?**
- **How can we improve the state-of-the-art for computer security experiments?**

What is an experiment?

- Experiment: A procedure in which an intervention is deliberately introduced to observe its effects.
- There are several types of experiment:
  - True experiment: random assignment to the treatment or alternative condition.
  - Quasi-experiment: not assigned randomly.
  - Natural experiment: Not really an experiment; the cause usually cannot be manipulated, e.g., in a study contrasting a naturally occurring event such as before and after an earthquake.
  - Correlational / observational experiment: a study that simply observes the size and direction of a relationship among variables.

True (randomized) experiment



What constitutes a GOOD experiment?

- Validity
  - Internal
  - External
- Control (of bias/error; eliminate alternative explanations)
- Repeatability
- Reliability
- Reporting (including *all* of the method)

- Asking the right questions

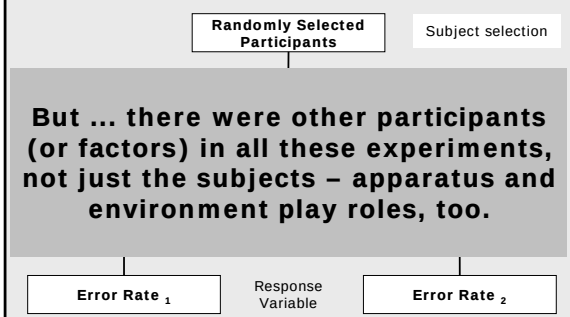
Example from keystroke dynamics

- First papers were published in 1978-1980.

- What question was asked?
  - Can you distinguish among users on the basis of their typing rhythms? Which classifier works best?
  - Typical experiment – N users type self-selected passwords; distinguish among users with classifier
  - After 30 years … answers are still unsatisfying.

- A different, perhaps more relevant, question …
  - Do people have unique typing rhythms?
  - Typical experiment – similar, but tightly controlled

True (randomized) experiment



A few influential factors …

- Different (and different-length) passwords
- Self-selected (not assigned) passwords
- No timing calibrations (one study -- 14% bad timestamps); resolution probably inadequate
- Different numbers of repetitions of passwords
- Noise from network, applications, timing, operating system, keyboard, logging
- Dropped subjects (questionable rationales)
- Practiced vs. unpracticed subjects (practice levels)
- Idiosyncratic or unknown outlier treatment

Results may be due to user typing rhythms, or to various other factors (same as intrusion detection)

Moral

- Security experiments can be good experiments, but they need to …
  - ask the right questions
  - be well designed
  - be valid
  - be repeatable
  - be generalizable
  - be explanatory
  - be reported thoroughly

- Otherwise, why bother?

Charge to the speakers …

- **What makes a good security experiment?**
- **What can and cannot be learned about security through experiments?**
- **Should there be better connections between formal and experimental security work?**
- **How can we improve the state-of-the-art for computer security experiments?**

State of the art ???

- First …
  - It's not the state of the art that's in trouble.
  - The state of the art is fine.
  - It's the state of the practice that's in trouble.
- Second …
  - If there's an art, it lies in asking good questions, and in devising valid experiments to answer them.
- But …
  - Perhaps we can improve by looking at current impediments to good experimentation … and removing or mitigating them.

Impediments (in no particular order)

- Community
  - There is no community collective that shares in common problems, methods, experiments and data, as in biology, medicine, epidemiology, cognitive science, physics, etc.
  - Communities are not supported - not as communities, and not as long-term research thrusts, with continuity.
  - Single laboratories can't do everything – invent the instruments, create the paradigms, run the experiments, do the analyses, etc. It's too much for one lab … especially in 18-36 months.
- Free and easy access to other research
  - Too much literature, too spread out, too hard to find, and too expensive
  - No public-access model, like NIH

Impediments (in no particular order)

- Incentive
  - Rewards are for novelty and silver bullets – shooting the moon
  - Few rewards for careful experimentation
  - No rewards for replication
  - Disincentives for careful and thorough reporting of methods
    - Although the culture seldom sees the need for thorough reporting anyway.
    - Note: the method is more important than the result
  - Referee community rejects as useless and boring
- Culture
  - The security culture does not embrace fully rigorous measurement and experimentation
  - They say they do, but when it comes down to it, they don't.
  - The culture rejects serious efforts as being too hard, the problems are too big, too many parameters, too complex, etc.
  - These are excuses; other fields have the same issues.
  - We may try, and fail, and try again; but not trying is failing.

Impediments (in no particular order)

- Infrastructure
  - Barriers to entry are high (too high)
  - We lack shared testbeds, experimental apparatus and experimental paradigms for gathering or generating data.
    - What about Geni, DETER, and NCR?
  - We lack shared benchmark data sets (with calibrated ground truth, and meta-data).
    - What about UNM, Darpa-98/99, Predict ?
  - We lack a shared analytical framework.
    - Shared tools, like R for statistics
    - Common scripts for data generation or handling
    - Common mechanism for exact replication of experiments

Impediments (in no particular order)

- Literacy
  - The community lacks the background and knowledge to conduct proper experimentation.
  - Unawareness of the fundamentals of experimentation, e.g., internal or external validity, control of confounds, elimination of alternative explanations, or experimental design.
  - There are few educational programs in experimentation.
- Wrong questions
  - Can we build a better gizmo … vs …
  - Why is the new gizmo better, and how does it generalize?
  - … or, what do the errors reveal?
  - We need insight, not just demonstrations.

What we need right now

- Support for community effort; continuity
- Shared benchmark data
- Shared methodologies
- Shared instrumentation
- Good scientific questions
- Good reporting practices in the literature; start with, at least, a complete methods section.
- Cooperative referees who won't dis good reporting
- Reproducibility/replicability
- Validity
- Operational definitions
- Education at the undergrad and grad levels; maybe corporate, too
- A shift in the culture

Charge to the speakers ... summary

- **What makes a good security experiment?**
  - **Look at what makes a good experiment.**
  - **Need education.**
- **What can and cannot be learned about security through experiments?**
  - **Depends on the questions being asked.**
- **Should there be better connections between formal and experimental security work?**
  - **Yes, of course.**
- **How can we improve the state-of-the-art for computer security experiments?**
  - **Remove or mitigate impediments.**
  - **Change the culture.**

- End – End – End – End – End – End -

XXX