Empirical Computer Security
ECE/CS 8803 ECS

Frank Li
To Do's for Next Class (8/24)

• Join Canvas + Piazza, look at the course website + syllabus. Let me know if anything is missing!

• Look over the class topics and think of the ones you might be interested in leading. I will be requesting your top 5 preferences via Canvas.

• Do we have brave volunteers for next week? Great papers, and I will help you :)

• **Homework:** Read Monday's paper and submit the paper summary by Sunday 5pm EST (paper summary questions are already posted).

• Look at my example paper summary responses for today's reading, posted.
Last Class: Traditional vs Empirical Security

Traditional Security:
- **Defensive**: For a given system/protocol, what are security properties we'd like, and how do we provide those properties (ideally w/ formal guarantees)
- **Offensive**: Given a system/protocol, how do we violate desired security properties

Empirical Security:
- Explore how security concerns manifest in practice
- Beyond considering a system/protocol in isolation/abstract, investigate how it is used in practice, how attackers actually attack it, and why.
- Insights inform defenses, including those beyond traditional technical ones
- Complementary to traditional security
Science of Security Paper


Cormac Herley
Microsoft Research, Redmond, WA, USA
cormac@microsoft.com

P.C. van Oorschot
Carleton University, Ottawa, ON, Canada
paulv@scs.carleton.ca

Abstract—The past ten years has seen increasing calls to make security research more "scientific". On the surface, most agree that this is desirable, given universal recognition of "science" as a positive force. However, we find that there is little clarity on what "scientific" means in the context of computer security research, or consensus on what a "Science of Security" should look like. We selectively review work in the history and philosophy of science and more recent work under the label "Science of Security". We explore what has been done under the theme of relating science and security, put this in context with historical science, and offer observations and insights we hope may motivate further exploration and guidance. Among our findings are that practices on which the rest of science has reached consensus appear little used or recognized in security, and a pattern of methodological errors continues unaddressed.

Index Terms—security research; science of security; history of science; philosophy of science; connections between research and research) in the light of consensus views of science and scientific methods. We find that aspects from the philosophy of science on which most other communities have reached consensus appear surprisingly little used in security, including in work done under the SoS label. For example, we do not find that that work better adheres to scientific principles than other security research in any readily identifiable way.

We identify several opportunities that may help drive security research forward in a more scientific fashion, and on this we are cautiously optimistic. While we see great benefit to this, we also do not wish to argue that all of security must be done on rigidly scientific principles. A significant component of security is engineering; this shares with science the regular contact with, and feedback from, observation, despite not

Note: Link to the paper on the course website
Two Sides of the Coin: Formalism

Security through formalism/proofs:

• Deductive logic: given a set of assumptions (or axioms), can reason and reach a logically certain conclusion about the security of a system

• Strengths? Weaknesses?
  • Strength: Diffie-Hellman key exchange cryptographic protocol is provably secure, assuming computing discrete logarithms are computationally hard.
  • Weakness: Side channel attacks (math world != real computers)
Two Sides of the Coin: Empiricism

Security through empiricism:

• Inductive logic: make observations and draw generalizations/conclusions from the data

• Strengths? Weaknesses?
  • Strengths: Password expirations results in worse password security
  • Weakness: Security on Twitter ≠ Security on other websites
Complementary Approaches

- Produce logically certain conclusions about security (assuming our assumptions hold)
- Inform and validate formal model assumptions and conclusions in reality

Formal Security ➔ Empirical Security
Good Empirical Methodology is Hard!

**Challenge 1:** Real-world data is messy and imperfect!
(Example: Clock precision & accuracy)

**Challenge 2:** We aren't perfect, we make mistakes. Are we really measuring what we think we are?
(Example: Proxies/caching for network measurements)
Good Empirical Methodology is Hard!

Management strategies:

Strategies for Sound Internet Measurement

Vern Paxson
International Computer Science Institute
Berkeley, CA 94704  USA
vern@icir.org

Note: Link to the paper on the course website
Good Empirical Methodology is Hard!

Management strategies:

• Perform extensive logging of your measurement itself (Example: periodic ping tests to identify network outages)
Good Empirical Methodology is Hard!

Management strategies:

• Perform extensive logging of your measurement itself (Example: periodic ping tests to identify network outages)

• Perform calibration:
  • Look at outliers/spikes

Figure 3: HTTPS Patch Rate. We track vulnerable web servers in the Alexa Top 1 Million and the public IPv4 address space. We
Good Empirical Methodology is Hard!

Management strategies:

• Perform extensive logging of your measurement itself (Example: periodic ping tests to identify network outages)

• Perform calibration:
  • Look at outliers/spikes

Figure 7: CDFs of CVE vulnerability life spans, for all CVEs and when grouped by software repositories.
Good Empirical Methodology is Hard!

Management strategies:

• Perform extensive logging of your measurement itself (Example: periodic ping tests to identify network outages)

• Perform calibration:
  • Look at outliers/spikes
  • **Employ self-consistency checks**
    (Example: # TCP SYN-ACKs <= # TCP SYN)
Good Empirical Methodology is Hard!

Management strategies:

• Perform extensive logging of your measurement itself (Example: periodic ping tests to identify network outages)

• Perform calibration:
  • Look at outliers/spikes
  • Employ self-consistency checks (Example: \# TCP SYN-ACKs \leq \# TCP SYN)

• **Measure multiple ways**
  (Example: measure from geographically diverse servers)
Good Empirical Methodology is Hard!

Management strategies:

• Perform extensive logging of your measurement itself (Example: periodic ping tests to identify network outages)

• Perform calibration:
  • Look at outliers/spikes
  • Employ self-consistency checks (Example: \# TCP SYN-ACKs \(\leq\) \# TCP SYN)
  • Measure multiple ways (Example: measure from geographically diverse servers)

• Test on known/synthetic data
Good Empirical Methodology is Hard!

**Challenge 3:** Lots of data
- Data squirrel (within reason)
- Start on small datasets first

**Challenge 4:** Reproducibility
- Ideally design reproducible methods/analyses
- Can greatly help you
- Also, science!
Good Empirical Methodology is Hard!

Challenge 5: Stats/math is hard :(  
- Multiple test corrections
Good Empirical Methodology is Hard!

Challenge 5: Stats/math is hard :

Let's say 1M emails are sent to GT folks each day. However, 100 are phishing emails. Is a phishing classifier with 0.1% false positive rate good?

Probably not! 0.1% FPR * 1M emails = 1K false positive emails a day! Classifier finds mostly non-phishing! (Base rate fallacy)

Or maybe? (Can you triage 1K FPs a day?)
Good Empirical Methodology is Hard!

CHALLENGE 6.1: ETHICS

Belmont Report on Ethics:

1. **Respect for persons:** protect participant autonomy, primarily through informed consent. Deception or no consent may be necessary and justified.
   - Good example: user study consent form
   - Bad example: undisclosed censorship measurements for website visitors

2. **Beneficence:** Do no harm, minimize risk.
   - Good example: crawling a website slowly
   - Bad example: vigilantism, testing leaked passwords, running malware uncontained

3. **Justice:** Equitable distribution of risk and benefits.
   - Good example: user study of developer security practices
   - Bad example: ?
Good Empirical Methodology is Hard!

**CHALLENGE 6.2: LAW**

Don't break the law...

Laws on hacking (Computer Fraud and Abuse Act - CFAA)

Laws on what communication/network traffic information you can intercept and store (Wiretap Act, Pen Register Act, Stored Communication Act)

Laws on user data collection, disclosure, and storage (Europe's General Data Protection Regulation - GDPR, and California Consumer Privacy Act - CCPA)

Laws protecting specialized populations (Children's Online Privacy Protection Rule - COPPA)

Legal case laws too! *(LinkedIn vs HiQ scraping case)*