Research, Generating Ideas, and Reading Literature

CSE545 - Fall 2006
Introduction Computer and Network Security
Professor McDaniel
What is research?

- Which activities are research?
  - Designing a new protocol?
  - Building an implementation of a protocol?
  - Measuring the cost of the protocol?
  - Formally evaluating the correctness of a protocol?
  - Developing methods of implementing, evaluation a protocol?
What is not research?

- Arguing the quality of a protocol?
- Arguing the appropriateness of a protocol?
- Surveying a field?
- Illustrating a limitation of a common practice or system?
A cynical definition:

• That which counts on your vita … is research.

• The hardest thing about a PhD is figuring out what “research” is …
Research vs. engineering

- Novelty …
- Importance … (sort of)
- Discovering a new fact or idea

- Engineering is often harder than research
- One must be careful to understand the difference
Research vs. Opinion

• Arguing a position is not research unless it uncovers some new thought or methodological device
  ‣ Difference is subtle

• Experts will often produce manifesto about an area
    – The key here is that they are *experts* and have the bona fides to make some an argument
    – This is *not research*
Why is there so much bad research?

• Most papers (90%+) I encounter are *bad* --- for one or more of the following reasons. The authors …
  … don’t formulate the problem well (or at all).
  … don’t motivate the problem well (or at all).
  … address an unimportant or moot problem.
  … are not familiar with the breadth or depth of the area.
  … do not discuss important related work.
  … realize the problem has been solved (or at least better addressed).
  … don’t have a coherent solution or it does not solve the problem.
  … don’t have a coherent or appropriate methodology.
  … don’t apply the methodology well.
  … don’t draw the correct conclusions from the results.
  … don’t present the work well enough to be understandable.
  … don’t articulate the take away.

• Any paper failing to do any of these things is a failure.
Security Research

• Almost as diverse as computer science itself
  ‣ Systems design
  ‣ Formal analysis
  ‣ Programming languages
  ‣ Hardware design
  ‣ Software engineering
  ‣ Human computer interfaces
  ‣ Networking, …

• Some are specific to security
  ‣ Cryptography
  ‣ Security protocol design
  ‣ Security Policy, …
How to read a paper ...

- Everyone has their own method for reading a paper
- However, you need to identify the following:
  - Area and motivation
  - Problem statement
  - Important related work
  - Solution
  - Methodology and assumptions
  - Main results
  - Take-away
  - Future work
A Review Exercise

• Form groups of mixed masters and PhD students (no more than 3 per group)
  • Everyone take a 15 minute scan of the paper
  • Write a 1 paragraph description of each of the 8 elements in previous page
  • Write a critique paper:
    • What is good about this paper?
    • What is bad/wrong/flawed in this paper?
    • 1 paragraph review of the paper?
Readings Critique

• Select one of the most important papers in your area appearing in IEEE Security and Privacy, ACM CCS, or USENIX Security (everyone must select a unique paper … check with people in your area)
  
  • Write a two page critique of the papers. First start with a discussion of the paper (see 8 element list in preceding slide)
    
    • Then, write a critique of the paper:
      
      • What is good about this paper (1/2 page)
      • What is bad/wrong/flawed in this paper (1/2 page)
      • 1 paragraph review of the paper (make qualitative judgment of the paper)
Idea Formulation

• The essential part of successful research is picking good problems and solutions?

The problem is real.
Idea Formulation (cont.)

- Good approaches to finding ideas:
  - First, read several papers (make sure they are good ones) in a particular area.
    - If this is a new topic area, you must become familiar with the problems, solutions, and terminology of the community.
  - The ask the the following questions (write down answers)
    - What are the problems that this area asks?
    - What are the methodological tools that people bring to bear in addressing problems in this area?
    - How is the field evolving?
    - How do your set of skills apply to the problems being addressed?
    - How are expected changes in the larger computer science community going to affect the known problems and solutions?

- Paper: “Patch on Demand” Saves Even More Time?
• Do the following exercises:
  ‣ (5 minutes) Listing: make a quick list of 1-5 word phrases that would be used by/related to/observance of the field and problems and solutions
    • This is not an outline, there is no ordering to the list
    • Use your imagination
    • Creativity is the essence of this exercise (don't overthink)
    • Some of list will be nonsense, do not filter thoughts
  ‣ Example: if I were looking at a paper about firewalls, I might come up with the following (just a start):
    • policy validation, distributed firewalls, bad for detecting viruses, …
    • Of course, this is general, should contain thoughts more specific to paper content,
    • e.g., better algorithm than Bob (the author) -- use graph theory
My results (from Keromytis paper)

- Virus, virus detection, automated patching, metamorphism, polymorphism, zero day vulnerability, metamorphism -- detection avoided by obscuring behavior, malicious behavior, automatic patching lowers availability, automatic patching makes knowing the security posture impossible (no idea what patches are on system), is zero day really the worst case?, vaccination system requires a precise definition of malicious behavior, honeypots, if you could detect it they why would you need to immunize, why does this make for better security?, seems very ad hoc -- nearly impossible to formulate anything concrete about how secure your system is, would I ever turn my network over to this thing?, what the hell is a remote sensor and why do I trust it, CCDC works great if you can talk to it -- DOS?, …
Using the results

• Examine closely the contents -- they will tell a story find singletons or clusters or phrases and see if they provide some new angle on a problem or issue

• For example, I choose: metamorphism -- detection avoided by obscuring behavior

• Which leads the following idea:
  ‣ Q: How do you make a better detector?
  ‣ Q: You look at behavior, not code.
  ‣ Who knows how to model the potential bahaviors of a program? Programming languages people: methodology: static or dynamic analysis

• Paper: “Detecting Metamorphing Viruses through Static Analysis”
Authorship

• This is the most dangerous part of publishing. This has led to the most serious rifts in the profession …
  • Make sure that anyone involved knows the policy (what one needs to do to be an author) the expectations and the repercussions of not participating as expected.
• Ordering matters in some fields (systems), not in others (math).
• Make sure everything is clear to everyone before getting started.
  • I have seen best friends never speak to each other again.
  • A paper is never worth that kind of heartache, but people will surprise you.
• Do you have a policy and what is it?
Idea Talks: 02/08/2007

• Everyone has 5 minutes to present idea (3 slides)
  ‣ Area
  ‣ Problem
  ‣ Related Works
  ‣ Solution
  ‣ Methodology
  ‣ Expected results
  ‣ Expected Take Away

• Everyone should practice timing and presentation. If the idea is not appropriate (tough love), or you don’t finish in 5 minutes, you will do it the next week.