Cybercrime Statistics Brainstorm Group

(designated volunteer: Sean Smith)

Outline

1. Standard arguments for what and why
2. Norms and definitions
3. Effective analytics, period
4. Privacy and the data
5. Resilience over time
6. Data for forensics and prosecution
7. The data is just the beginning
8. Meta-issues

1. Standard Arguments

- better measurements help with evaluation of how good we're doing
- tweak parameters and see what happens

- ties directly to HUMAN harm---and human causes
    - including better education for users!

- correctly measuring harm
- correct methodology to measure
- correct data to measure
- align incentives so it happens

1. Standard Arguments (continued)

- Whose data
  - Victims
  - Criminals
  - And law enforcement

- Compare with crime stats in physical world
  - Helpful?
  - Trouble from bad stats in the physical world
  - And bad stats in cyber!
    - FBI 1000x
    - False positives
    - Discovering when you start looking
- But also *distinguish* from physical crime stats

1. Standard Arguments (continued)

- Ancillary domains
  - National security
  - Risk managements
  - Insurance

- GC: Urgency: lost history
  - No knowledge of history
  - Data sources with longer-term horizons

2. Norms and definitions

- GC: Norms of what is a crime
  - What makes a crime “cyber”?
- GC: Or should we just be looking at cyber behavior, period?
  - To discover dependencies on non-crime factors
  - Which runs into norms again:
    - Privacy
    - Civil rights
- Countries interpret behaviors and evidence differently

3. GC: Effective analytics, period

- The wave is not going in that direction :(
- GC: Predictive analytics!
  - Trends
  - Changes in environments
- GC: Better data!
  - Insulate collection against pressures of profit, etc.
  - One avenue: outsourcing enterprise/agency email etc. to cloud providers
- GC: Understanding risk shifts
  - Adversaries are adaptive
  - Adversaries can be consumers of these stats too
  - (And some criminal service providers sell them!)

4. Privacy and this data

- The measurements themselves can violate privacy norms
- GC: How to balance privacy of the data vs. ability to do research on this data
- GC: Effective re-contribution of seized data
    - Credit card story
- “Center for Disclosure Avoidance Research” at US Census Bureau

5. Resilience across timescales

- We don’t want to merely measure technical artifacts
- Victimization rates
- GC: But challenge: what about the continuing reinvention of the technology?
    - E.g.: what crimes were suddenly made possible by social networking?
    - Example approach: computer-aided money laundering
        - It was valuable to do initial exploration in jurisdictions that didn’t have wiretap restrictions.

6. Forensics and Prosecution

- GC: How can we harness this data for effective prosecution?
- GC: Attribution
    - Including nation-state issues
- GC: Rules and procedures for “software on the witness stand”
- GC: What crime is not prosecuted?
    - GC: Is the paradigm of prosecution even scalable?
    - We need data to know!

7. The Data is just the beginning

- OK, so we might have good data about who is exploiting what vulns where.
- GC: Why do these vulns even exist?
    - CVE vs CWE argument.
- GC: Why are these vulns being exploited and not others?
- GCL Why are these sites/places/countries being attacked, and not others?

8. Meta issues

- Barriers? Addressed above
- Incremental progress is indeed possible
- In the physical world, it took many decades!

Illustration (not exhaustive)