

Metricizing (Mis)Configuration

Roy A. Maxion

Dependable Systems Laboratory
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213

Email: maxion@cs.cmu.edu

11-12 August 2008
Workshop on Assurable and Usable Security Configuration
National Science Foundation
Fairfax, Virginia

Metricizing

- To metricize is to make metrical; i.e., relating to measurement.
- Why metricize (mis)configuration and its contributing factors?
 - Because we are trying to be scientific.
 - Because science is nothing if it is not metrical.
- Measurement is the most fundamental method in science;
it is the process of empirical, objective assignment of numbers (measurements) to the properties of objects and events in the real world in such a way as to describe them.

L. Finkelstein, *Theory and Philosophy of Measurement*, 1982.

Metricizing

- "When you can measure what you are speaking about, and express it in numbers, you know something about it; but **when you cannot measure it**, when you cannot express it in numbers, **your knowledge is of a meager and unsatisfactory kind**: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science." -- Lord Kelvin, Popular Lectures and Addresses [1891-1894].

Science

- Science is the systematic effort to **discover** (and to **understand**) how things work.
- Science uses controlled methods ... usually experiments ... to observe natural phenomena ... to **explain** ... in a **reproducible** way.
- Science makes claims, and supports them with **credible evidence**.
- A touchstone of science is to **predict** ... and to **generalize** from specific findings to broader, often future, situations.
- Without measurement, all this will be ... a bit hard.

Today ...

- Discuss a few illustrative details, drawn largely from other areas (e.g., keystroke dynamics), but equally important in usable security and privacy, and in configuration management.
 - Clear goals
 - Operational definitions
 - Instrumentation and calibration
 - Instructions to subjects (and experimenters)
 - Sampling
 - Demographics
 - Experimental methods

A few factors influencing experiments

- Target goal (presence or absence)
- Operational definitions
- Sampling
- Calibration of measurement instruments
- Instructions to experimenters/participants
- Demographics
- Experimental method

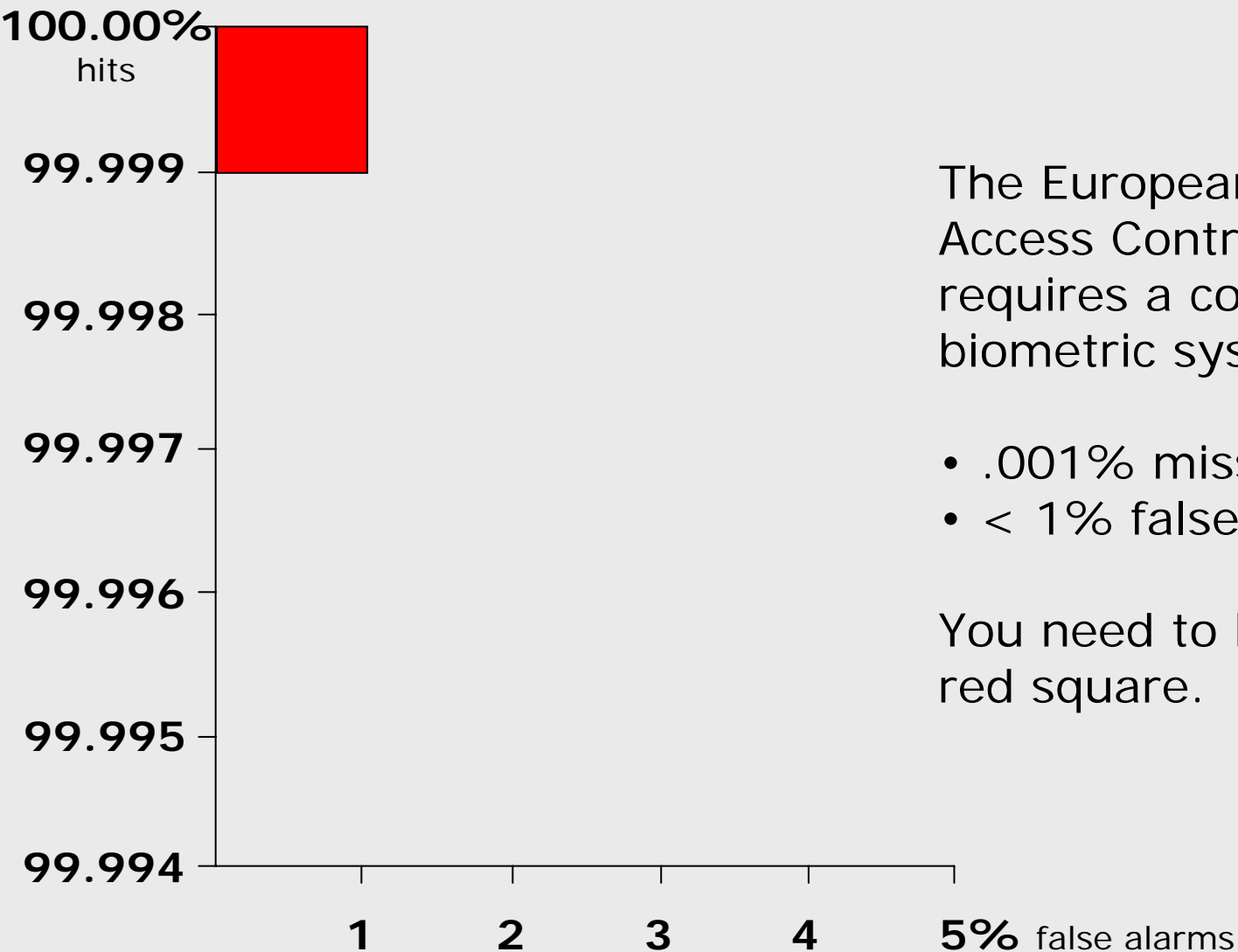
Target goal

- What is the (measurable) target?
- If you don't know, you won't hit it.
- You won't know how far you are from it.
- You won't know what to do to reach it.

- Keystrokes:
 - Five 9s of classification accuracy
 - 99.999% accurate

- Usable security/configuration:
 - ???
 - 100% of novice users will correctly set file permissions in a complex, high-stress setting.

Target: how good do you have to be?



The European Standard for Access Control (EN 50133-1) requires a commercial biometric system to have a

- .001% miss rate; and
- < 1% false alarm rate.

You need to be in the little red square.

Operational definition

- A precise, unambiguous description of how to obtain a repeatable value for a characteristic being measured. It includes a precise definition of the characteristic, and how to measure it.

Used to remove ambiguity and ensure all data collectors have the same understanding.

- Examples

- A recipe (for food) is an operational definition of the dish being prepared (e.g., PB&J sandwich).
- How tall is a school child?

- Questions

- What does “usable” mean? How is it measured?
- What does “configuration” mean? How is it measured (especially when it’s only partially correct)?
- What is a measure of assurance?

Instrumentation and calibration

- We often use apparatus and measuring instruments to gather data.
- We may forget that these instruments are subject to uncertainties of their own.
- It is essential to obtain credible, valid, repeatable measurements from our instruments; otherwise, error contorts our results.
- Examples
 - Keystroke data - noise
 - Survey data (even demographic surveys are instruments)
 - Usability and correct-configuration data

Instructions to subjects

- When people participate in an experiment, they're told what to do.
- If the instructions are ambiguous (or wrong), you won't get the behavior you're trying to investigate.
- It's critical to report instructions in your results; omissions can mislead.
- Examples:
 - Typing a password – it's not a contest
 - Configuring a system to do a certain thing

Sampling

- Since testing an entire population is generally not feasible, we rely on sub-populations – samples.
- Samples must be representative of the general population about which we wish to draw conclusions and make predictions.
- You cannot generalize beyond your sample.
- Examples:
 - 1936 election: Landon to win, Roosevelt to lose
 - Convenience and snowball samples
 - Skill level (typing or configuring) – low, high?

Demographics

- Learning details about your subjects (or other experimental materials) can be essential for understanding outcomes.
- In the keystroke domain ...
 - Handedness – left or right?
 - How did you learn to type – school or self?
 - Do you touch-type, or hunt-and-peck?
 - Other ... (e.g., neural issues)
- In the security domain ...
 - How configuration knowledge was acquired
 - Whether the subject is expert or novice
 - Cognitive preferences for problem-solving types

Experimental method

- The method matters ... a lot.
- A flawed method can make a good experiment look bad, and a bad experiment look good. (Blackburn et al. 2008)
- Most important – eliminate biases and confounds.
- Examples from the keystroke domain ...
 - Choosing your own password
 - Different keyboards
 - Self-selecting web pages in mouse dynamics
- Examples from the security domain ...
 - Mixing skill levels or cognitive preferences
 - Mixing hard and easy configuration problems
 - Unawareness of subject's motivation level

Experimental method – serious stuff

FBI forensics – chemical analysis of bullet lead

- FBI metallurgist was uncomfortable; no listeners.
- He had a few questions ...
 - Is the chemical profile of a batch really unique?
 - How large would a batch have to be before specificity becomes meaningless?
 - What is the right statistical test for assessing a match?
 - Are analytical procedures consistently followed?
- A few answers ...
 - 27% false-alarm rate (35 million bullets in a batch)
 - In an outdoor-oriented town like Juneau (pop. 30k), most people would be suspects
 - Now inadmissible under Federal Rules of Evidence
- FBI no longer uses bullet-lead analysis – too flawed

Federal rules of evidence (rule 702)

- Testimony must ...
 - be based upon sufficient facts or data
 - be the product of reliable principles and methods
 - have applied the principles and methods reliably to the facts of the case.
- The key for the Court in determining credibility of evidence is primarily one of...

"reliability of method."

How good is our science?

- Shouldn't our science be at least as good as legal evidence requires? Is it now?
- Is it ...
 - Reproducible
 - Falsifiable
 - Valid, not confounded, not biased
 - Predictive
 - Methodologically sound
 - Measurable

Summation

(Tony Tether speaking about the NCR)

- Over the ages, scientific progress has been held back by the ability to make measurements at the level of the environment for which the scientific research was being done ... telescopes, microscopes, particle accelerators, etc.
- The National Cyber Range is the measurement capability for cyber research in both classified and unclassified environments. Without it [measurement], research will be done in darkness and only stumble accidentally into the light.
- ... and only stumble accidentally into the light.
- We need good measurement & good methodology.

- End – End – End – End – End – End -

