Accountability and Deterrence in Online Life

Joan Feigenbaum
http://www.cs.yale.edu/homes/jf/

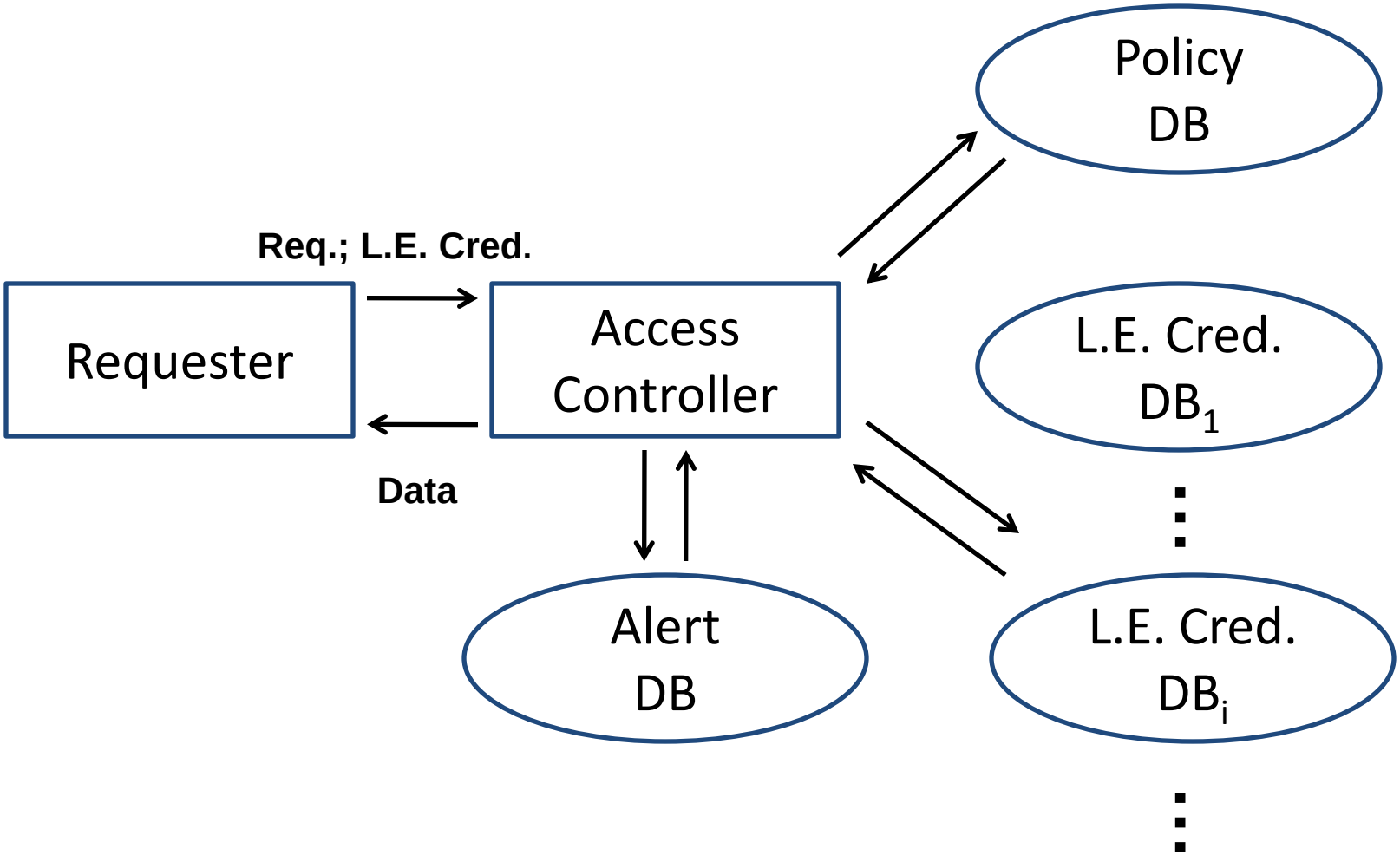
3rd Web Sciences Conf.; June 16, 2011

Joint work with James A. Hendler, Aaron D. Jaggard, Daniel J. Weitzner, and Rebecca N. Wright

Utopian Vision of IT Security

- Design and implement IT systems in which users ***cannot*** break the rules.
- Use ***preventive mechanisms***, *e.g.*:
  - Passwords
  - Authentication protocols
  - Digital signatures
- Before a user can take a security-sensitive action, he must ***prove that he is authorized*** to do so. Attempts at, *e.g.*, "breaking into" a network, forging a digital signature, or eavesdropping on encrypted communication, should be technically infeasible.
- No need for police, courts, and lawyers!

Law-enforcement officials (and ONLY they) may access the Alert Database.



“©2011, Disney. All rights reserved.”

- DRM systems allow only authorized users to access the content and restricts the manner in which they can use it.
- Under the Fair-Use provisions of US copyright law, certain categories of uses do not require authorization by the rights holder.

? A user may need to ***access the work in order to determine how he wants to use it*** (and thus whether he needs authorization).

Our Thesis

The preventive approach to security and privacy is increasingly inadequate in online life.

Accountability mechanisms are needed to complement preventive mechanisms.

"When a policy-governed action occurs, it should be possible to determine (perhaps after the fact) whether an applicable policy has been violated and, if so, to have the ***violators face appropriate consequences***."

Our Contributions

- Realistic scenarios that support the need for accountability mechanisms
- Comparison of existing formal frameworks for accountability
- Terminological issue: Is “accountability” the best word for our purpose?

Examples of Scenarios in which Prevention is Inadequate

- Surveillance
- Emergencies (“break-glass” scenarios)
- Data exchange in (evolving!) online life
  - Search, e-commerce, and social networking
  - P2P photo sharing
  - Location-dependent services
- High-volume transactions
  - robots.txt
  - Consumer finance

Research Goal: Define “Accountability”

- There is widespread agreement that “accountability” is important in online activity, but people disagree about what it means.
- Users will resist the construction of a “cyber architecture for accountability” if they think its cost (in, *e.g.*, privacy, speed, or convenience) will be too high.
- Progress on definitions and terminology may defuse this resistance and identify *desiderata* for architectures and protocols.

A Difficult Concept to Nail Down!

- “Accountability is a protean concept, a placeholder for multiple contemporary anxieties.” [Mashaw, 2005 – administrative law]
- “[A]ccountability has not yet had time to accumulate a substantial tradition of academic analysis. ... [T]here has been little agreement, or even common ground of disagreement, over the general nature of accountability or its various mechanisms.” [Mulgan, 2003 – political science]

Defn. of Grant & Keohane, 2005

- Accountability exists in global-scale inter-actions when some actors have the right to
  - hold other actors to a set of standards,
  - judge whether they have fulfilled their responsibilities in light of these standards, and
  - impose sanctions if they have not.
- G & K assume an official international framework; a country's unilateral defense of its interests is *not* an accountability mechanism.

Prior Work in Comp. Sci. (1)

- "Accountability is the ability to hold an entity, such as a person or organization, responsible for its actions." [Lampson, 2005 – 2009]

- "An accountability protocol gives [an agent] lasting evidence, typically digitally signed, about the actions performed by his peer." [Bella & Paulson, 2006]

Prior Work in Comp. Sci. (2)

- Applications of Lampson's definition
  - Accountable Internet Protocol [Andersen *et al.*, 2008]
  - Social-web applications [MIT Decentralized Inf. Group]
- Cryptographic applications in which participants remain anonymous unless they break the rules
  - Electronic cash
  - Anon. group messaging [Corrigan-Gibbs & Ford, 2010]

Limitations of Prior Approaches

- Reliance on identification (or at least persistent identities) of those held accountable
- Reliance on an authority to hold an entity accountable
- Need for the authority to take an explicit action in order to hold an entity accountable

New Framework [FJW, 2011]

An entity *i* is accountable for obeying policy P if, whenever *i* violates *P*, then, with some positive probability, *i* could be punished.

- Entities' actions represented as "event traces"
- Separates accountability from identifiability
- Relaxes Lampson's definition by allowing *automatic* (or "passive") punishment
- Punishment ≈ expected utility is decreased
- Decreased wrt what? A "normal" protocol trace (as used in [Halpern, 2008] to study causality)

Automatic Punishment without Identifiability [Vickrey, 1961]

2^{nd}-price auctions: Policy is "Bid your true value."

- For many natural distributions on the bidders' values, a bidder cannot improve his utility by lying; indeed, with positive probability, his utility will be decreased if he lies about his value.
- No punishing action is taken; thus, this is *automatic* punishment
- The violator isn't identified!
- Nobody else even knows that there was a violation!!

Mediated Punishment with Partial (or no) Identifiability [Lampson, 2009]

Policy: Don't send spam.

Accountability mechanism: "Reject email unless it is signed by someone you know or comes with 'optional postage' in the form of a link certified by a third party you trust, such as Amazon or the U.S. Postal Service; if you click the link, the sender contributes a dollar to a charity."

Deterring Spam, cont.

- The punishing action is taken by the receiver and the trusted third party.
- The receiver need not know the sender's identity; the third party may or may not need to know it, depending on the payment system.
- Sender *i*'s utility for trace T is $\omega_i(T) - q$
  - $\omega_i(T)$ = dollar value that i assigns to T
  - q = # msgs deemed by receivers to be spam

Terminological Question

- Will users resist "accountability mechanisms" in online life because, in common parlance (unlike in FJW), the term seems to preclude anonymity and suggests actively "accounting for oneself" to an authority figure?
- Recall that our end goal is to ***deter violations*** by having ***violators face appropriate consequences***.
- Is "deterrence" a better term than "accountability" for this very general notion?

Ongoing and Future Work (1)

- Explicate relationships among basic S&P concepts in online life:
  - Authorization
  - Deterrence
  - Anonymity and pseudonymy
  - Detection
  - Identification
  - Accountability
  - Punishment
  - Compensation

Ongoing and Future Work (2)

- Enforcement!
- Formal analysis of up-and-running accountability mechanisms that show promise experimentally, *e.g.*:
  - P2P currencies such as “iOwe” [Levin *et al.*, NetEcon 2011]
  - “PeerReview” in distributed systems [Haeberlen *et al.*, SOSP 2007]