Accountability in Computing

Joan Feigenbaum

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

Northwestern IDEAL Institute workshop; April 30, 2021

What is “Accountability”?

- “Accountability is a protean concept – a placeholder for multiple contemporary anxieties.”

  Jerry L. Mashaw, Professor Emeritus of administrative law

  Yale Law School

- “Accountability is a core concept of public administration, yet disagreement about its meaning is masked by consensus on its importance and desirability.”

  Jonathan G. S. Koppell

  Dean of the College Of Public Service & Community Solutions

  Arizona State University

Talk Outline

- Motivation
- Highlights from a recent survey*
- One approach to “accountability” as conceived by NSF’s Law and Science program**

---

* J. Feigenbaum, A. D. Jaggard, and R. N. Wright, *Accountability in Computing: Concepts and Mechanisms*, Foundations and Trends in Privacy and Security 2(4) (2020), pp. 247–399.

** Designing Accountable Software Systems, https://www.nsf.gov/pubs/2021/nsf21554/nsf21554.htm

Motivation for the Study of “Accountability” in Computing over the last 20+ Years

- Adoption of Internet-scale, policy-governed systems
  - Examples: Social-media platforms with “community-standards” policies
  - Traditional “preventive” approach to security, privacy, and authorization is no longer adequate.
    - Users are numerous, diverse, and scattered; information about them is scarce.
    - Access and authorization decisions are no longer binary.
    - Policies are dynamic and require timely information.
  - Alternative approach: Hold users accountable for policy violations.
- Proliferation of laws and regulations about information and systems
  - Examples: GDPR, CCPA, proposed modifications of Section 230
  - How can system developers be held accountable to legal requirements?

Even Simple Formulations Can Be Subtle (1)

- Lampson (2005): Accountability is the ability to hold an entity, such as a person or organization, responsible for its actions.
- Grant and Keohane (2005): Accountability implies that some actors have the right to hold other actors to a set of standards, to judge whether they have fulfilled their responsibilities in light of these standards, and to impose sanctions if they determine that these responsibilities have not been met.

Even Simple Formulations Can Be Subtle (2)

- Lampson (2005): Accountability is the ability to hold an entity, such as a person or organization, responsible for its actions.
- Grant and Keohane (2005): Accountability implies that some actors have the right to hold other actors to a set of standards, to judge whether they have fulfilled their responsibilities in light of these standards, and to impose sanctions if they determine that these responsibilities have not been met.
- "Rights" are a central focus in political science.
  - An entity might have the right to do something but not the ability to do it.
- CS focuses on technical capabilities and limitations of system entities.
  - Why care about the "right" to do something that one is not technical able to do?

Even Simple Formulations Can Be Subtle (3)

- Lampson (2005): Accountability is the ability to hold an entity, such as a person or organization, responsible for its actions.
- Grant and Keohane (2005): Accountability implies that some actors have the right to hold other actors to a set of standards, to judge whether they have fulfilled their responsibilities in light of these standards, and to impose sanctions if they determine that these responsibilities have not been met.
- Lampson doesn't say judgment & sanctions are done by the same entity.
  - In fact, he doesn't say anything at all about judgment or sanctions.
- "Accountability" is a *system property* in Lampson's formulation.
  - *How* entities are held responsible is not specified.

Spectrum of Accountability-Related Activities

- **Prevention:** Plays a role before a policy violation occurs
- **Violation**: Plays a role at the time a violation occurs
- **Detection:** Facilitates, enables, *etc.*, discovery of a violation either at the time the violation occurs or afterward
- **Evidence:** Gathers or preserves evidence about a violation that may be used against an accused violator. Can play a role before, during, or afterward
- **Judgment:** Renders a verdict about an actor's guilt or blameworthiness with respect to a violation. Plays a role after a violation occurs
- **Punishment:** Penalizes a violator after a violation occurs

- A single accountability mechanism might be involved at multiple points on the spectrum.

Key Questions about Accountability Definitions and Mechanisms Include:

- Do users of the policy-governed system have persistent IDs?
  - Is a particular notion of "accountability" consistent with anonymity or pseudonymy?
- Is the mechanism centralized or decentralized?
  - Does it respond to a violation (in gathering evidence, judging, or punishing) in a centralized or decentralized fashion?
  - Is strategic behavior by administratively independent parties in a decentralized system a potential obstacle to achieving accountability?
- Must evidence of a violation be presented to a judge?
  - Is the judge a participant in the system or external to it?
- Is punishment automatic, or is it imposed by a designated party?
  - Is the punishing party a participant in the system or external to it?

Example: Accountable Internet Protocol (1)

Andersen *et al.* (2008)

- Definition: Accountability is the association of an action with the responsible entity.
  - Applications of such associations include
    - Prevention and detection of source-address spoofing
    - Stopping of unwanted traffic
  - Such associations work for both violating and non-violating entities.
- Core technique: Self-certifying addresses
  - The name (*e.g.*, the host ID) of an object is the hash of the public key that corresponds to that object (*e.g.*, the hash of the host's public key).
  - Cryptographic-signature verification can be used, *e.g.*, to ensure that the only packets that are forwarded are those with correct (unspoofed) source addresses.

Example: Accountable Internet Protocol (2)

Andersen *et al.* (2008)

- Focuses on the *prevention* and *detection* points on the spectrum
- *Requires* persistent IDs
- *Decentralized* (like almost everything in the Internet)
- To the extent that there is punishment, it is *imposed internally* by network participants who implement AIP's "shut-off protocol."

Example: PeerReview (1)

Haeberlen *et al.* (2007)

- Definition: An accountable system is one that maintains a tamper-evident record that provides non-repudiable evidence of all nodes' actions.
- Core techniques and features:
  - Identities: Each action is undeniably linked to the node that performed it.
  - Secure record: The system maintains a record of past actions such that nodes cannot secretly omit, falsify, or tamper with its entries.
  - Auditing: The secure record can be inspected for signs of faults.
  - Evidence: When an auditor detects a fault, it can obtain evidence of the fault that can be verified independently by a third party.

Example: PeerReview (2)

Haeberlen *et al.* (2007)

- Focuses on the *detection*, *evidence,* and *judgment* points on the spectrum
- *Requires* persistent IDs
- *Decentralized*
- The evidence gathered must be able to convince an *external* party that a violation has been committed.
- Although PeerReview *does not include a punishment function*, Haeberlen et al. (2007) states that one of PeerReview's benefits is "deterrence," which it provides through "threat of punishment."

Example: Feigenbaum, Jaggard, & Wright (2011) (1)

- Definition: An entity is accountable for obeying a policy if, whenever it violates the policy, it can be punished. When punishment occurs it must be a result of the violation.
- Core techniques:
  - Utility functions
  - Event traces
  - Reasoning about causality (as in Lamport (1978), Halpern (2008), *etc.*)
- Participants in an accountable, policy-governed system have utilities that change as events occur.
  - A violator can wind up with a lower utility than it would have had if it had obeyed the policy.
  - Its loss of utility is *caused by* the violation (not, say, by "bad luck").

Example: Feigenbaum, Jaggard, & Wright (2011) (2)

- Focuses exclusively on the *punishment* point on the spectrum. Rationale is that, without punishment, the mechanism has *enabled* accountability but has not actually held the violator accountable.
- FJW11 framework is otherwise fully general.
  - Persistent IDs, temporary IDs, anonymity, pseudonymy, …
  - Centralized or decentralized
  - The mechanism may or may not require that evidence be presented to a judge. The judge can be internal or external to the accountable system.
  - Punishment can be automatic or administered by a designated party. That party can be internal or external.
- Econ notion of incentive compatibility satisfies the FJW definition.
  - Consider an online-auction system in which the policy is “bid your true value,” and truthfulness is a dominant strategy.
  - A bidder who violates the policy (and only such a bidder) may lower his utility.

Objection: Weitzner (2017)

- That definition is too general.
  - A violator might be punished without having been identified and judged.
  - The violator himself might not even understand that he has violated the policy or know that he has been punished.
  - This framework is devoid of the interactive, social, and educational role that accountability mechanisms typically play in communities.
- What (FJW, 2011) have defined is *deterrence*, not accountability.
- "One sense of 'accountability', on which all are agreed, is that associated with the process of being called 'to account' to some authority for one's actions." [Mulgan (2000)]

Approach/Paper	Time/Goals: Prevention	Detection	Evidence	Judgment/Blame	Punishment
Internal Evidence (Sec. 3.1)					
AIP	(✓)	✓			(Med.)
APIP	(✓)	✓			(Med.)
PGPA		✓	(✓)		
Packet passports	(✓)		✓		(Med.)
AudIt/packet obit.		✓			
Evidence for Third Parties (Sec. 3.2)					
CATS		✓	✓	(✓)	
Accountable-subgroup multisig.	✓	✓	(✓)		
PeerReview & AVMs		✓	✓	(✓)	
Cryptographic commitments		✓	✓		
Time stamping		✓	✓		
Judgment or Blame (Sec. 3.3)					
DISSENT		✓	✓	✓	
Jagadeesan *et al.*, 2009				✓	
Barth *et al.*, 2007		✓	(✓)	✓	
Punishment (Sec. 3.4)					
A2SOCs			✓	✓	(Med.)
CHL off-line e-Cash		✓	✓	✓	Med.
B–LB Reputation				(✓)	Med.
PEREA					Med.
iOwe		✓	✓		Med.
Non-equivocation contracts		(✓)	(✓)		Med.

Table 3.1: Time-and-goals of accountability systems and mechanisms.

	Information		
Approach/Paper	Identity Requirements for Participation	Violation Disclosed?	Violator Identified as Such?
Internal Evidence (Sec. 3.1)			
AIP	Host	Broad	Broad
APIP	Unique	Broad	Unique
PGPA	(Unique)	Limited	Limited
Packet passports	Key	Limited	Limited
AudIt/packet obit.	Broad	Unique/Limited	Unique
Evidence for Third Parties (Sec. 3.2)			
CATS	Key	Broad	Broad
Accountable-subgp. multisig.	Broad	Unique/Broad	No/Broad
PeerReview & AVMs	Broad	Broad	Broad
Crypto. commitments			
Time stamping	Key	Limited	Limited
Judgment or Blame (Sec. 3.3)			
DISSENT	Key	Broad	Broad
Jagadeesan *et al.*, 2009	(Broad)	(Limited)	Unique
Barth *et al.*, 2007	(Broad)	Unique	Unique
Punishment (Sec. 3.4)			
A2SOCs	Unique	Broad	Broad
CHL off-line e-Cash	Key	Broad	Broad
B–LB Reputation	Broad	Broad	Broad
PEREA	Key	Unique	No
iOwe	Key	Broad	Broad
Non-equiv. contracts	Key	Broad	Broad

Table 3.2: Information classification of accountability systems and mechanisms.

	Action			
Approach/Paper	Centralization without Violation?	Centralization with Violation?	Punishing Entity?	Requires Ongoing Involvement?
Internal Evidence (Sec. 3.1)				
AIP	Dec.	Dec.	Int.	Yes
APIP	Dec.	Dec.	Int.	Yes
PGPA	Dec.	Dec.		
Packet passports	Dec.	Dec.	Int.	Yes
AudIt/packet obit.	Dec.	Dec.	Ext.	No
Evidence for Third Parties (Sec. 3.2)				
CATS	Dec.	Dec.	No	
Accountable-subgp. multisig.	Dec.	Dec.		
PeerReview & AVMs	Dec.	Dec.		
Crypto. commitments				
Time stamping	Cent.	Cent.		
Judgment or Blame (Sec. 3.3)				
DISSENT	Dec.	Dec.	No	
Jagadeesan *et al.*, 2009	Dec.	Dec.	No	No
Barth *et al.*, 2007	Cent.	Cent.	No	
Punishment (Sec. 3.4)				
A2SOCs	Cent.	Cent.	(Int.)	(Yes)
CHL off-line e-Cash	Cent.	Cent.	Int.	No
B–LB Reputation	Dec.	Dec.	Int.	Yes
PEREA	Dec.	Dec.	Int.	No
iOwe	Dec.	Dec.	Int.	Yes
Non-equiv. contracts	Dec.	Dec.	Int.	No

Table 3.3: Action classification of accountability systems and mechanisms.

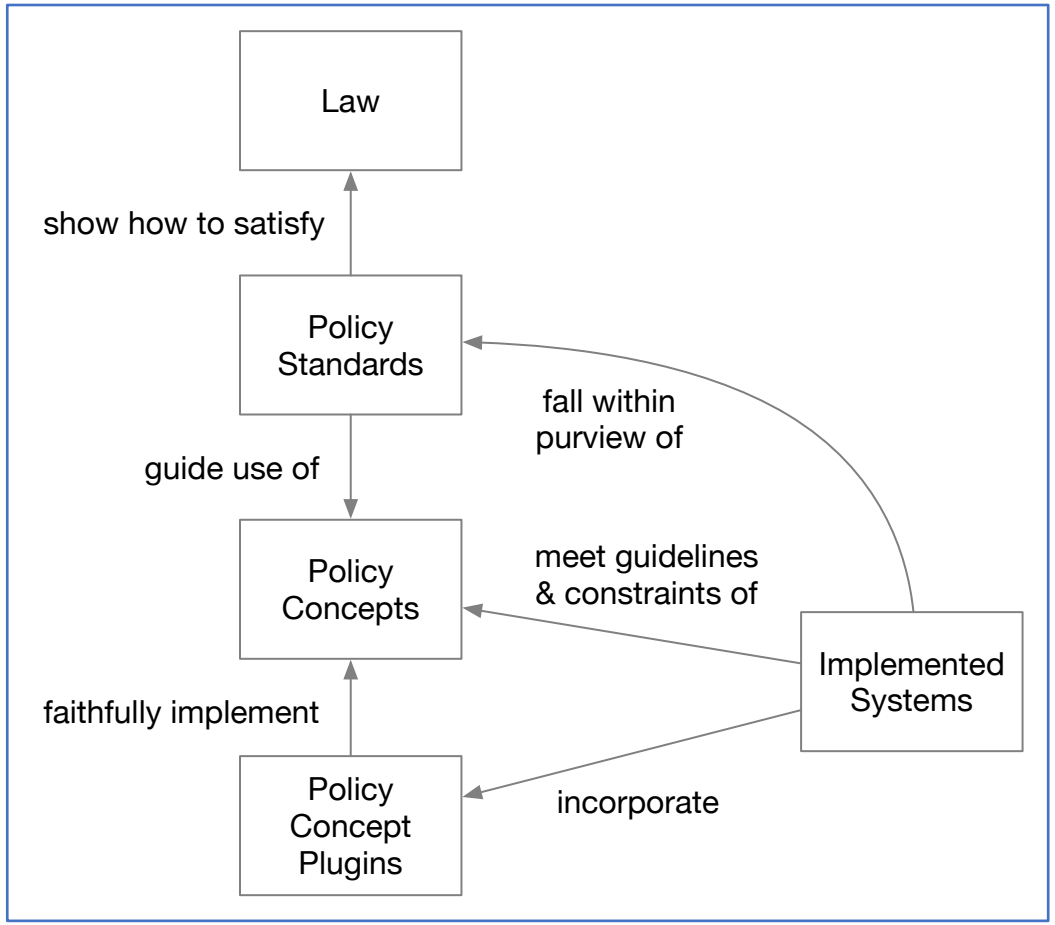
Designing Accountable Software Systems (DASS): NSF (2021)

- Earlier work focused on holding participants in a system accountable to the system policy.
- DASS program shifts the focus to holding designers and implementers of software systems accountable to legal requirements.
- Feigenbaum, Jackson, and Weitzner (2021): There is a "double accountability gap."
  - Gap between the actual and intended policy-relevant behavior of software components. Software developers need better, policy-aware tools.
  - Gap between legislative language and the software-system behavior that the law is actually supposed to require. Legislators need tools for the complex, interpretive task of expressing policy constraints that can be implemented in software.
- Software engineering to the rescue!

Legally Accountable Cryptographic Computing Systems (LAChS)

- Policy concepts
  - High-level, rigorously described software-design patterns
  - Identify the functional aspects of software systems in order to assess whether they are consistent with the policy constraints
- Policy standards
  - Functional descriptions of the requirements of law
- Policy soundness
  - Conceptual and logical connections between legal requirements and software artifacts
  - Enables formal reasoning about the soundness of a software artifact with respect to a provision of law
  - Proof of the policy soundness of a software system is confirmation that it is accountable to legal requirements

LAChS Approach to Achieving Policy Soundness



Some “Accountability in Computing” Milestones

- Nissenbaum, 1997
  - First paper to foreground the word “accountability” in the study of computer systems
  - Inserted accountability into the discourse on human values in computers and software
- Weitzner *et al.*, 2008; Lampson, 2009
  - Brought accountability to a prominent position in the study of online privacy
  - Emphasized the inadequacy of preventive privacy technology in Internet-scale computing
- Feigenbaum, Jaggard, and Wright, 2011
  - Shifted the focus to “punishment”: Tie violating actions to consequences
  - Decouple from identification. Is “accountability” different from “deterrence”?
- Kroll, 2015; Kroll *et al.*, 2017; Frankel *et al.*, 2018
  - Concluded that accountability mechanisms are essential in an era of mass surveillance
  - Proposed accountability mechanisms make essential use of cryptographic computing
- National Science Foundation, 2021
  - NSF Solicitation 21-554: Designing Accountable Software Systems (DASS)

Questions?