Accountability and Deterrence in Online Life (Extended Abstract)

Joan Feigenbaum
Yale University – CS Dept.
P.O. Box 208285
New Haven, CT 06520-8285
(203) 432-6432
joan.feigenbaum@yale.edu

James A. Hendler
RPI – CS and Cog. Sci. Depts.
110 8th Street
Troy, NY 12180-3590
(518) 276-4401
hendler@cs.rpi.edu

Aaron D. Jaggard
Colgate University – CS Dept.
13 Oak Drive
Hamilton, NY 13346-1398
(315) 228-7650
adj@dimacs.rutgers.edu

Daniel J. Weitzner
MIT – CSAIL
32 Vassar Street, Room 32G-524
Cambridge, MA 02139-4309
(617) 253-5702
djweitzner@csail.mit.edu

Rebecca N. Wright
Rutgers University – CS Dept. and DIMACS
96 Frelinghuysen Road
Piscataway, NJ 08854-8018
(732) 445-5931
rebecca.wright@rutgers.edu

ABSTRACT

The standard technical approach to privacy and security in online life is *preventive*: Before someone can access confidential data or take any other action that implicates privacy or security, he should be required to prove that he is authorized to do so. As the scale and complexity of online activity has grown, it has become apparent that the preventive approach is inadequate; thus, a growing set of information-security researchers has embraced greater reliance on *accountability mechanisms* to complement preventive measures. Despite widespread agreement that "accountability" is important in online life, the term has no standard definition. We make three contributions to the study of accountability: (1) We flesh out with realistic examples our claim that a purely preventive approach to security is inadequate; (2) We present, compare, and contrast some existing formal frameworks for accountability; (3) We explore the question of whether "deterrence" may be a better general term in this context than "accountability."

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues— privacy

General Terms

Economics, Security, Legal Aspects

Keywords

Accountability, Deterrence, Privacy, Security

1. INTRODUCTION

The utopian dream of many cryptography and security researchers has long been a digital world in which people are unable to break the rules. In such a world, encryption, authentication, digital signatures, and other security mechanisms make it technically infeasible for people to read others' confidential communication, access others' computers and networks, distribute others' copyright material, *etc*., without permission. Thus, the basic technical approach to online privacy and security has been a *preventive* one: Before someone can take an action that is governed by a privacy or security policy, he is required to prove that he is authorized to take it.

In online life, which is characterized by Internet commerce, social networking, web-accessible health records, personalized search, and many other ways to engage socially, economically, and intellectually with both strangers and friends, preventive mechanisms are increasingly inadequate. As a result, a growing faction in the cryptography and security community (see, *e.g.*, [15, 24]) has embraced greater reliance on *accountability* mechanisms to complement preventive measures: 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. A move in this direction would make the online world more like the offline world, in which potential violations of security and privacy are often deterred by the prospect of negative consequences rather than prevented by truly unbreakable locks.

Despite widespread agreement that "accountability" is important in online life, the term has no standard definition. In this work, we make three contributions to the study of accountability: Sec. 2 fleshes out with realistic examples our claim that a purely preventive approach to security is inadequate; Sec. 3 presents, compares, and contrasts some existing formal frameworks for

accountability, including a very general framework put forth in recent work by three of us [7]; Sec. 4 explores the question of whether "deterrence" may be a better general term in this context than "accountability." Some open problems are discussed briefly in Sec. 5.

2. THE NEED FOR ACCOUNTABILITY MECHANISMS

In this section, we provide realistic examples to support our claim that a purely preventive approach to security and privacy is inadequate. We first consider scenarios in which a potential user must gain access to sensitive or valuable information in order to determine whether he is entitled to use it and, if so, for what purposes. We then consider emergency situations, in which rules that are strictly enforced under typical conditions may need to be bent or broken, but these exceptions cannot simply be ignored once the emergency is over. Next, we point out that online life is characterized by data exchanges in which notions of "appropriate use" are evolving rapidly and are highly unlikely to be captured any time soon by logically complete policies. Finally, we note that there are environments in which purely preventive approaches could work in principle but, in practice, would impose undesirably high computational costs. These examples are illustrative and are not intended as an exhaustive list of scenarios in which accountability mechanisms can be useful. For a more in-depth discussion of the need for accountability mechanisms, see, *e.g.,* [6, 15, 24].

2.1 Data-Dependent Notions of Appropriate Use

US copyright law provides a crisp example of a set of policies that cannot be enforced in a purely preventive manner if they are to achieve their goals. In general, copyright law specifies that an author (or, more generally, the creator of a copyright work) has certain exclusive rights, including the right to control copying and distribution of his work and the right to authorize or refuse to authorize the creation of derivative works (such as sequels or movie versions of books). However, there are important exceptions to these rights that are encoded in the *fair-use* provisions of copyright law. A researcher, for example, may, under the fair-use doctrine, make a small number of copies of a scientific journal paper for use by his research group without obtaining the author's permission, but he may not (without the author's permission) share the article with everyone at his university or some other wide audience. A properly attributed excerpt from a newspaper story may, without the author's permission, be copied and distributed widely without infringing copyright or committing plagiarism. The notion of fair use promotes socially desirable activities, such as education and criticism, and is regarded by many as an essential pillar of cultural production.

In the analog world of printed books, journals, and newspapers, there is no attempt at preventive enforcement of copyright law. It is technologically feasible to make and distribute a large number of unauthorized copies of a book, thus violating copyright law; if one does so, however, one runs the risk of being caught and sued for copyright infringement (*i.e.*, being "held accountable" for one's illegal actions), and in any case one incurs the non-negligible cost of copying and distribution. The fact that copyright enforcement is based on detection rather than prevention helps enable fair use: In order to determine how, if at all, one wants to use a document and whether such a use requires the author's permission, one must be able to read (and, in particular, to have access to) the document; preventive copyright enforcement might restrict access to those who can first justify that access or pay for it.

The flowering of digital copyright works and online distribution has brought with it attempts at preventive copyright enforcement. Digital-Rights-Management (DRM) systems are justified in part by the negligible cost of copying and distributing digital works. Unfortunately, DRM systems often subject *all* users to strict limits (or even prohibition) on uncompensated access to the works they manage. It is very difficult to design these limits in a manner that is consistent with the goals of copyright law; if the limits are very strict, they threaten fair use, but, if they are too permissive, the works might be too easily copied and distributed and the creators' rights vitiated. The digital-copyright problem is rendered more complex by the prevalence online of "mashups," in which pieces of different documents, often with different protections or subject to different fair-use rules, are combined and redistributed.

It is our thesis that access and accountability together form a better approach to digital copyright than draconian forms of preventive DRM. Allowing users to access copyright works online, just as they access analog works when browsing in physical stores and libraries, is consistent with preservation of the creators' rights provided that they are held accountable for subsequent use of those works in accordance with copyright law.

Similar challenges are present in surveillance: A law-enforcement officer may not be able to determine whether he needs a warrant in order to tap a data stream, because the nationalities of sender and receiver cannot be determined without access to the very data that he wants to tap. As in the digital-copyright scenario, more flexible and effective surveillance-policy enforcement can be achieved through temporary access and after-the-fact accountability than through attempts to prevent violations from ever occurring.

2.2 Emergency Use

There are a number of common emergency scenarios in which there is a clear need to augment or complement traditional, preventive access controls and usage policies. They are often called *break-glass* scenarios, in order to conjure up the image of breaking a glass to pull a fire alarm (see, *e.g.,* [11]).

Consider, for example, the case of a physician in one medical practice, who is not, as a routine matter, allowed to access the patient records of another medical practice. If that physician were in the presence of a patient of the other practice who needs emergency assistance, she could present her medical ID to the other practice, together with a brief description of the nature of the emergency, and be granted temporary access to the patient's records. The description and her ID number should be logged, and all emergency-access logs should be audited periodically. The legal and professional repercussions of being caught using one's medical ID in this manner when not in a *bona fide* emergency would deter abuse of the emergency-access system and thus support patients' privacy – in other words, physicians would be "held accountable" for proper emergency use of patients' records.

As another example, consider military information-classification systems, which often include rules that are roughly tantamount to "A *top-secret* document may not be accessed by a person who has only a *secret* clearance – unless the latter's life is in danger, and the sharing is authorized by a user with the appropriate *top-secret* clearance." In current military IT systems, exercising such an

exception may force a user with secret clearance to work around the security system, because "user's life is in danger" is not something the access-control system can verify automatically, and the user's "secret" authorization credentials will cause the system to deny him access to the top-secret document. In an accountability-based system, a remote user with top-secret clearance could grant temporary access to the user with secret clearance, but he would get a warning from the security system that such access is exceptional and be required to provide a brief explanation of why he believes that a life is in danger. As in the medical scenario, these exceptional accesses and accompanying justifications would be logged and audited.

2.3 Security and Privacy in an Evolving Online World

In the activities that currently dominate online life, including Web search, shopping, and social networking, users transmit a great deal of sensitive, personal information about themselves, both to Web-based companies and to other users. Exactly what the companies and individuals who receive this information should be allowed to use it for and how long they should be allowed to retain it are hotly debated issues. Until there are laws or at least widely agreed-upon social norms governing "appropriate use," it will be impossible to devise logically complete, machine-readable policies, let alone to use them in a preventive fashion to avoid misuse. It may be feasible, however, to keep track of how sensitive information is used, to analyze these uses and devise "best practices" in an iterative fashion, and to hold accountable parties that use it in clearly objectionable ways.

Privacy policies have received a great deal of media attention as major companies such as Google, Facebook, and Twitter periodically change their policies and defaults (especially when introducing new services and features), sometimes in significant ways, and sometimes in ways that greatly confuse users. Although it is easy in hindsight to criticize some of the decisions that these companies have made, it is actually quite difficult to design usable privacy technology and good defaults. An accountability framework that enabled users to determine who has accessed their information, how it has spread around the network, and what it has been used for could increase their understanding of the impact of their privacy settings and help them learn (albeit by trial and error) to make better choices.

Accountability mechanisms could also be used to enhance privacy options and to give users more flexibility, interactivity, and personalization than they currently have online. User A may send a photo to user B with certain restrictions, *e.g.*, "please don't post this online." If B were to forget and start to upload the photo to a website, he could receive a warning such as "you agreed not to do this; if you continue, the person who sent you this photo will be informed." In some cases, B might decide to continue (perhaps thinking "I'm posting to a local, well managed site that I know A would approve of"), and in others he might not. With appropriate mechanisms, these instructions could be propagated, and a third user C who received the photo would, upon starting to post it, discover that A did not want it to be widely shared. The fact that "don't post this online" is actually just an approximation of A's policy for the photo (chosen either because she does not know precisely how she wants the photo to be used or because she cannot express her wishes precisely in the relevant policy language) is not overly restrictive when use is governed by an accountability system, but it could be overly restrictive if use were governed by a preventive access-control system.

2.4 Computational Efficiency of Accountability Mechanisms

Finally, note that there are scenarios in which prevention could work in principle but would in practice impose undesirably high computational costs. One basic example is the "robots.txt" protocol, which is used on the Web to stop search engines from indexing certain pages. Rather than a heavyweight access-control mechanism, robots.txt is essentially a social protocol – a web server provides to the search engine a document that specifies in machine-readable form what should and should not be indexed. There is no enforcement mechanism. However, the web server's access logs make clear which crawlers have touched which pages. If a search engine crawls a site and indexes the pages that it shouldn't, it can be held accountable. The web server can take a technical step (barring that search engine from the site altogether) or a social one – when reputable search engines violate the robots.txt protocol, it makes the news.

The same approach is used in finance, where signatures on small checks often go unverified unless and until a dispute arises. Also in the realm of commercial transactions, note that, when you order a book from Amazon.com, you get an email message stating that the purchase has occurred and providing a means to complain if you believe that an error has been made. This works at the granularity of a book but would be unbearable at the granularity of a micropayment in a multi-user game. For the game, it makes more sense to set a bound on what one thinks is appropriate and get warnings only if the total of one's micropayments exceeds the bound. This approach is already found in the modern web (consider, *e.g.,* 3G plans in which a user pays each time he goes over a pre-set limit but not for each download), but there is not yet a formal theory that tells us what protections it can and can't provide.

3. FORMALIZING "ACCOUNTABILITY"

In this section, we consider definitions of accountability that appear in the literature (both in Computer Science and in other fields) and evaluate their applicability to online life. Three illustrative (but far from exhaustive) examples are the definitions put forth by Grant and Keohane [8], Lampson [14], and Feigenbaum *et al.* [7]. We then discuss the relationship of these definitions to other approaches to accountability that do not explicitly define the term. Finally, we provide an overview of the formal framework used by Feigenbaum *et al.* to capture the notion of punishment and, in turn, that of accountability.

3.1 Defining Accountability

The notion of accountability predates the development of "online life." Yet, even before it became an issue in online life, accountability was difficult to clarify. This is recognized by those who have studied accountability in the context of administrative law or public policy: Mashaw [16] notes that "accountability is a protean concept, a placeholder for multiple contemporary anxieties," while Mulgan [18] observes that "accountability has not yet had time to accumulate a substantial tradition of academic analysis" and that "there has been little agreement, or even common ground of disagreement, over the general nature of accountability or its various mechanisms."

Mulgan [17] argues that what is embodied in "accountability" has expanded, "los[ing] some of its former straightforwardness and [coming] to require constant clarification and increasingly complex categorization." He states that a core meaning of accountability is and has been related to "the process of being

called 'to account' to some authority for one's actions" (citing [12]), and he identifies features of the core notion of accountability: It involves giving account to an entity *external* to the accountable entity; it involves *social interaction and exchange* between the accountable entity and the entity calling for the account; and it involves the *rights of authority* that belong to the entity calling for the account. Mulgan also provides a survey of ways in which "accountability" has been expanded in recent decades beyond these core features. These new aspects include "responsibility," an internal (instead of purely external) aspect, "control," "responsiveness," and "dialogue." Interestingly, Dubnick [5] argues that accountability is strongly tied to the English language and English history, with many other major languages either lacking a direct counterpart entirely or translating only a narrower concept.

Mulgan [17] notes that "the inclusion of sanctions in the core of accountability is contestable on the grounds that it may appear to go beyond the notion of 'giving an account.' On the other hand, '*calling* to account,' as commonly understood, appears incomplete without a process of rectification." Below, we will focus on definitions of accountability that include some aspect of sanctions, "holding responsible," or punishment.

Grant and Keohane focus on the interaction of nation states and define accountability as the "right of some actors 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 independently puts forth a similar definition in a technical context: "Accountability is the ability to hold an entity, such as a person or organization, responsible for its actions." Although sensible and useful, these definitions do not address the rich spectrum of need for accountability in online life. Feigenbaum *et al.* call an entity "accountable with respect to a policy" if, whenever the entity violates the policy, then with some positive probability it is, or could be, punished.

It is worth distinguishing between these definitions. In addition to focusing on inter-state relations and assuming multinational frameworks, Grant and Keohane frame their definition in terms of the *right* to do various things when responsibilities are not met. The *ability* to do these things may be inherent in the setting they consider; for accountability in online life, however, it seems useful to have ability as an *explicit* part of the definition of accountability. Indeed, at least when defining punishment, which is a building block for the definition of accountability in [7], it may be preferable not to require the right to do something that might serve as punishment. Inflicting punishment without the right to do so may itself be a punishable violation of some standard, but definitions for online life should nevertheless be broad enough to encompass this.

The definition of Feigenbaum *et al.* differs from Lampson's definition particularly in its explicit focus on (potentially passively) being punished instead of on (actively) inflicting punishment, which seems a natural reading of "hold[ing] ... responsible." Importantly, and as discussed below, this definition intentionally does not require that a violator be identified, only that he be punishable. The fact that a violation occurred need not even be known (and an example in which it is not know will be presented in Sec. 3.3 below); as discussed in [7], this may foster the decoupling of identity from accountability in online scenarios. It also contrasts with the idea of "hold[ing] an entity ... responsible" insofar as the latter suggests knowledge of the identity of the entity being held responsible. Note that the definition in [7] does not require that the violator actually *be* punished. In particular scenarios (*e.g.*, an ongoing law-enforcement investigation that might be jeopardized by punishing a minor offence), there may be other reasons not to punish a violator; such a decision should not mean that the violator is not "accountable" for the violation. The framework of Feigenbaum *et al.* distinguishes an entity's "being accountable" from "being accountable to another entity," the second of which is used to indicate that a violator could indeed be punished and that this would be done in an active – "mediated" – way by that other entity.

Although these definitions are similar in spirit, we thus see that there are various subtleties that need to be addressed in considering accountability in online life. Additionally, full formalization (for use in analyzing an online or other system) of an accountability definition like one of these requires a formal definition of when the violator has had sanctions imposed upon him, been held responsible, or been punished. We discuss the approach of [7] to these questions in Sec. 3.3.

3.2 Other Approaches to Accountability

The study of systems for accountability in public administration has been a topic of interest; as but one example, Romzek and Dubnick [21] systematically studied accountability systems in the public sector, identifying *bureaucratic*, *legal*, *professional*, and *political* systems. Considering the types of accountability systems defined in his work with Romzek, Dubnick [5] subsequently framed an "ethical theory" approach to accountability; this built on the ethical theory paradigm of Nozick [20], which includes internal "moral pushes" and external "moral pulls." He then explored what accountability would mean in these different types of systems. As one example, Dubnick considers legal settings with external (legal) liabilities and argues that "accountability will be more whole and effective" when legal behavior results from a sense of internal obligation – the moral push – instead of the desire to avoid legal liability – the moral pull.

More closely related to the problems of interest here, there have also been various approaches to accountability in online life. They fall on different points of a spectrum that includes the gathering and presentation of evidence (of a violation), a judgment that a violation has occurred, and punishment for the violation. For example, in a typical legal process, all of these elements are present and occur in this order. Many protocols to provide some sort of "accountability" in online life are concerned with evidence and judgment. (See, *e.g.*, electronic-cash protocols [2, 3] and frameworks in which auditors blame agents [10], or judging agents deliver verdicts [13]. Similarly, Bella and Paulson [1], while not explicitly defining accountability, say that the "accountability protocols" they study give a participant "lasting evidence, typically digitally signed, about actions performed by his peer.")

The definitions of accountability that we focus on here frame accountability in terms of the possibility of punishment for violations of standards or policies. Importantly, these are orthogonal to particular methods of gathering and presenting evidence, judgment processes, and methods of punishment. Such questions are more context-specific; however, they provide many interesting open questions. As noted by Feigenbaum *et al.* and discussed in Sec. 3.3 below, the focus on punishment instead of on points earlier in this spectrum (*e.g.*, providing evidence that a particular entity violated a policy) raises the possibility of decoupling accountability from identity. This might allow

accountability and deterrence without simultaneously imposing identity systems that are not currently part of online life.

3.3 A Formal Model of Accountability that Focuses on Punishment

We now return to the formal accountability framework given by Feigenbaum *et al.* [7]. After they define accountability as described in Sec. 3.1, the main technical thrust of their work is the formalization of the notion of punishment in order to make the definition of accountability precise; integral elements of their formalism include event traces and utility functions. Essentially, they say that an entity (which we will speak about as if it were an individual, but which may be a group, a state, a company, or any other type of agent in a system) is punished for a violation if he is worse off – in some fashion – than if he had not committed the violation. They treat, in a unified manner, scenarios in which accountability is enforced automatically and those in which enforcement must be mediated by an authority; their framework includes scenarios in which the parties who are held accountable can remain anonymous and those in which they must be identified by the authorities to whom they are accountable.

Feigenbaum *et al.* [7] treat activity within a system (such as a computer network with access controls or a retail store with anti-shoplifting measures) as a sequence of abstract events (a trace); the value of a participant's utility function on a sequence of events indicates how much benefit the participant derives from that particular sequence of actions in the system. Using this approach, the authors formalize both "mediated" and "automatic" notions of punishment. Mediated punishment corresponds to what might come to mind most often when thinking about "accountability" and "punishment" – the fact that a violation occurred leads to a punishing action (which "mediates" the punishment) that makes the violator worse off than he would have been had the violation not been committed. By contrast, if a violation is automatically punished, then no mediating action is needed; the violator is worse off (again, in some sense, which might not be universal) immediately after the violation is committed. Both of these approaches have details that merit further discussion.

A violator should not be considered punished simply because, since the violation occurred, his utility has decreased; the decrease might be the result of something completely unrelated to the violation, the simplest example of which is bad luck. Similarly, an unrelated windfall might mean that the violator's utility is higher than when he committed the violation even though the violator might have been punished in an intuitive sense. For example, someone who shoplifts a book, wins the lottery, and then pays a fine for shoplifting that is much smaller than his lottery winnings has still been punished.

Automatic punishment is simpler to formalize than mediated punishment, but it lies outside of what we often think of as "holding someone accountable," because it does not involve a premeditated act of punishment. Second-price Vickrey auctions [23] serve as an intuitive motivating example: Imagine that the policy that might be violated is "bid your true value in the auction." Assuming a non-vanishing distribution on the other bidders' true values, it is well known that a bidder is worse off (probabilistically) if he violates this policy than if he bids his true value. However, even if the bids are such that violator is indeed worse off than if he had bid his true value (*e.g.*, if he wound up paying more than his true value for the good), the violator's identity (as a violator) is not revealed to the other participants; moreover, nobody even knows that the policy was violated.

In mediated punishment, the important question of how to determine what constitutes "worse off" is potentially subtle because of the events that may have taken place since the violation occurred; in particular, what is the reference value to which a violator's utility should be compared when deciding whether he has been punished for the violation? Another consideration is the connection of the punishing act to the violation itself – intuitively, if a violator is punished for one violation, we do not wish to consider him punished for all of his prior misdeeds as well. For example, someone might shoplift a book and then, a year later, steal a car. If he is then sentenced to five years in prison for stealing the car, we do not want to think of him as also having been punished for shoplifting even though he is then likely worse off than he was before he took the book. Finally, there may be many ways that events in a system may continue to unfold after some sequence of events has taken place, and the utility function of a policy violator may be unknown to those attempting to ensure that violators are punished. For both the different ways that a system may continue to evolve and the different possible utility functions, Feigenbaum *et al.* consider both probability distributions and "typicality rankings"; the latter indicate, *e.g.*, which utility function is most common in a population or which future scenario is most likely, but they do not assign probabilities to these. The different combinations of these approaches give different formal definitions that are appropriate for different models.

As suggested by the second-price auction example above, deterrence through automatic punishment may resemble the game-theoretic notion of incentive compatibility. When there are distributions (instead of ranking functions) on both the utilities and the future events in the system, the definition of automatic punishment in [7] coincides with *ex ante* Bayesian-Nash incentive compatibility. Relatedly, those trying to deter violations do not have knowledge of the potential violator's utility function, although the potential violator would naturally have this information when deciding whether to commit the violation. As a result, standard (interim) Bayesian-Nash incentive compatibility is more likely the ideal – but not always practical – solution concept, but that is distinct from what is provided by automatic punishment.

In formalizing mediated punishment, Feigenbaum *et al.* addressed the effects of unrelated events by comparing the violator's utility after the punishment with what his utility would have been if the violation and all subsequent events that were caused (in a formal sense) by the violation were omitted. For example, imagine the following sequence of events occurs: Alice buys a book from the bookstore; Bob shoplifts a rare book from the bookstore; Bob crashes his car into his mailbox; Alice (without knowing the book is stolen) pays Bob a $5,000 for the rare book that he stole; Bob wins a $10,000 prize in the lottery. Intuitively, if Bob is now going to be punished (in a mediated fashion) for shoplifting the book, his utility must – as a result of the punishment – be less than his utility after the sequence obtained by removing his violation and everything that depends on it, *i.e.*: Alice buys a book; Bob crashes his car; and Bob wins a $10,000 lottery prize. As noted above, Bob's utility might be viewed in either a typical or probabilistic fashion, and this might be computed with respect to either the typical or the (probabilistically) expected future evolution of the system.

Feigenbaum *et al.* address the issue of connecting the punishment to the violation (in the mediated case) by requiring that the punishing action be causally related to the fact that the violation occurred. In the preceding example, imposing a fine of $25,000

on Bob whenever Alice buys a book from the bookstore would not be a punishment for Bob's shoplifting, even though this presumably has the requisite effect on Bob's utility. On the other hand, imposing a fine on Bob as a result of a conviction in court, which in turn depended on evidence of Bob's shoplifting, would likely qualify as a punishing action. Implicit in this is the use of a formal approach to causal connections; the work of Halpern [9] on causality is used in [7].

4. ON TERMINOLOGY

Some recent work on formalizing "accountability," including [7], strives for the ability to capture everything not adequately handled by traditional preventive security mechanisms. This very general use of the word "accountability" may create barriers to adoption: In common parlance, the word connotes "standing up to be counted" in ways that suggest formal adjudication and the inability to act anonymously or pseudonymously while remaining accountable. In the security community, the term "accountable anonymity" is used, *e.g.*, in [4], to mean that a participant in a communication protocol remains anonymous unless he breaks the rules and disrupts the communication, at which point his identity is revealed; implicit in this notion is that loss of anonymity is what enables accountability. Feigenbaum *et al.* [7] emphasize the fact that their notions of punishment do *not* necessarily entail identification, but potential users of accountability mechanisms may find this counterintuitive and resist adoption. For this reason, we encourage the security-research community to consider the question of whether "deterrence," a word used by Lampson [15], is a better term for the most general notion and, if so, which forms of deterrence should be called accountability mechanisms.

As we explained in Sec. 3.1, Mulgan [17] identifies an apparent distinction between sanctions and the core of the "giving an account" view of accountability. In turn, deterrent effects may similarly be viewed as outside of the core definition of accountability, at least from a public-administration perspective. This adds at least some weight to the argument for using "deterrence" instead of "accountability" in these settings.

It does not seem that we can simply use an existing economic notion of incentive compatibility. As noted in Sec. 3.3, standard Bayesian-Nash incentive compatibility may capture an ideal notion of deterrence in models where it can be applied. However, this is much too strong a notion to require and thus not the right term to use. Furthermore, because there may not *be* a probability distribution on violators, *ex ante* incentive compatibility will not always work as a stand-in term.

5. CONCLUSION AND OPEN PROBLEMS

There is a growing body of work on accountability now available for reflection and analysis, including but not limited to the work discussed in Sec. 3. We regard it as the foundation for a more fulsome and formal understanding of accountability in online life and for the design, deployment, and use of accountable systems.

Moving from analysis and understanding to design, deployment, and use will require substantial technological progress. Today's technological infrastructure cannot straightforwardly support some of the accountability mechanisms that we have described. For example, it is conceptually simple to require that users who access copyright material be held accountable for the uses that they subsequently make of it, but there is currently no straightforward way to enforce such a requirement. Watermarks and other techniques that publishers have used to track digital works online have often been circumvented.

Ironically, some technological approaches might lead to acute tension between accountability and the very privacy and control that it is supposed to confer upon users. No one wants to have his every mouse click recorded and examined for compliance with a privacy policy, but it is unclear how to prove compliance without such a record.

As we discussed in Sec. 2 above, notions of appropriateness in Web search, social networking, and other activities of online life are evolving rapidly, making it infeasible to use standard preventive security measures such as authentication and encryption to enforce appropriate behavior. Although we believe that it would be more productive to strive for accountability than prevention when approaching this problem, accountability mechanisms by themselves will not solve it. Sloan and Warren [22] have considered this aspect of accountability, focusing on the challenges of applying the accountability model to privacy. They point out that, even with accountability mechanisms in place, there would still be many barriers to online privacy, including insufficient tools for representing privacy policies in machine-readable form, lack of context-sensitive techniques for reasoning about privacy (an issue that has also been explored in depth by Nissenbaum [19]), lack of widely shared social norms about online privacy, and inadequate social and legal incentives to maintain accountable systems.

6. ACKNOWLEDGMENTS

Professor Feigenbaum's research was supported in part by NSF grant CNS-1016875 and DARPA contract N66001-11-C-4018.

Professor Hendler's research was supported in part by NSF grant CNS-0831442 and IARPA grant FA8750-07-0037. This research is continuing through his participation in the Network Science Collaborative Technology Alliance sponsored by the U.S. Army Research Laboratory under Agreement Number W911NF-09-2-0053.

Professor Jaggard's research was supported in part by NSF grant CNS-1018557.

Mr. Weitzner's research was supported in part by NSF grant CNS-0831442 and IARPA grant FA8750-07-0031. Weitzner is currently serving as Deputy Chief Technology Officer for Internet Policy in the White House, but the ideas expressed in this paper are entirely his own and those of his coauthors; these ideas do not reflect the view of the Obama Administration.

Professor Wright's research was supported in part by NSF grant CNS-1018557.

7. REFERENCES

[1] Bella, G. and L. Paulson, "Accountability Protocols: Formalized and Verified," *ACM Transactions on Information and System Security*, vol. 9, no. 2, 2006, pp. 138 – 161.

[2] Camenisch, J., A. Lysyanskaya, and M. Meyerovich, "Endorsed E-Cash," in *Proceedings of the 28th IEEE Symposium on Security and Privacy*, 2007, pp. 101 – 115.

[3] Chaum, D. "Blind signatures for untraceable payments," in **CRYPTO '82**, Plenum Press, 1982, pp. 199 – 203.

[4] Corrigan-Gibbs, H. and B. Ford, ""Dissent: accountable anonymous group messaging," in *Proceedings of the 17th ACM Conference on Computer and Communication Security*, 2010, pp. 340 – 350.

[5] Dubnick, M. J. "Clarifying Accountability: An Ethical Theory Framework," in **Public Sector Ethics: Finding and**

Implementing Values, C. Sampford, N. Preston, and C.-A. Bois (eds.), The Federation Press, 1998, pp. 68 – 81.

[6] Feigenbaum, J. "Accountability as a Driver of Innovative Privacy Solutions," in *Privacy and Innovation Symposium*, Yale Law School Information Society Project, October 2010. http://www.law.yale.edu/intellectuallife/Privacy%20Symposium%20Thought%20Pieces.htm

[7] Feigenbaum, J., A. D. Jaggard, and R. N. Wright, "Towards a Formal Model of Accountability," submitted, April 2011.

[8] Grant, R. and R. Keohane, "Accountability and Abuses of Power in World Politics," *American Political Science Review*, vol. 99, no. 1, 2005, pp. 29 – 43.

[9] Halpern, J. "Defaults and Normality in Causal Structures," in *Proceedings of the 11th Conference on Principles of Knowledge Representation and Reasoning*, 2008, pp. 198 – 208.

[10] Jagadeesan, R., A. Jeffrey, C. Pitcher, and J. Riely, "Towards a Theory of Accountability and Audit," in *Proceedings of the 14th European Symposium on Research in Computer Security*, Lecture Notes in Computer Science, vol. 5789, Springer, Berlin, 2009, pp. 152 – 167.

[11] Joint NEMA/COCIR/JIRA Security and Privacy Committee (SPC), "Break-Glass – An Approach to Granting Emergency Access to Healthcare Systems," 2004, http://www.medicalimaging.org/wp-content/uploads/2011/02/Break-Glass_-_Emergency_Access_to_Healthcare_Systems.pdf

[12] Jones, G. W. "The search for local accountability," in **Strengthening Local Government in the 1990s**, S. Leach (ed.), Longman, 1992, pp. 49 – 78.

[13] Küsters, R., T. Truderung, and A. Vogt, "Accountability: Definition and Relationship to Verifiability," in *Proceedings of the 17th ACM Conference on Computer and Communications Security*, 2010, pp. 526 – 535.

[14] Lampson, B. Notes for presentation entitled "Accountability and Freedom," http://research.microsoft.com/en-us/um/people/blampson/slides/AccountabilityAndFreedom.ppt

[15] Lampson, B. "Usable Security: How to Get it," *Communications of the ACM*, vol. 52, no. 11, November 2009, pp. 25 – 27.

[16] Mashaw, J. "Structuring a Dense Complexity: Accountability and the Project of Administrative Law," *Issues in Legal Scholarship*, The Reformation of American Administrative Law, Article 4, 2005. http://www.bepress.com/ils/iss6/art4

[17] Mulgan, R. " 'Accountability': An Ever-Expanding Concept?," *Public Administration*, vol. 78, no. 3, 2000, pp. 555 – 573.

[18] Mulgan, R. **Holding Power to Account: Accountability in Modern Democracies**, Palgrave MacMillan, 2003.

[19] Nissenbaum, N. **Privacy in Context: Technology, Policy, and the Integrity of Social Life**, Stanford University Press, 2010.

[20] Nozick, R. **Philosophical Explanations**, Harvard University Press, 1981.

[21] Romzek, B. S. and M. J. Dubnick, "Accountability in the Public Sector: Lessons from the Challenger Tragedy," *Public Administration Review*, vol. 47, 1987, pp. 227 – 238.

[22] Sloan, R. H. and R. Warner, "Developing Foundations for Accountability Systems: Informational Norms and Context-Sensitive Judgments," in *Proceedings of the ACM Workshop on Governance of Technology, Information, and Policies*, 2010, pp. 21 – 26.

[23] Vickrey, W. "Counterspeculation, auctions, and competitive sealed tenders," *Journal of Finance*, vol. 16, no. 1, 1961, pp. 8 – 37.

[24] Weitzner, D. J., H. Abelson, T. Berners-Lee, J. Feigenbaum, J. Hendler, and G. Sussman, "Information Accountability," *Communications of the ACM*, vol. 51, no. 6, June 2008, pp. 82 – 88.