Forgive and Forget: Return to Obscurity

Matt Bishop
UC Davis
Davis, CA
mabishop@ucdavis.edu

Emily Rine Butler
University of Oregon
Eugene, OR
erbutler@uoregon.edu

Kevin Butler
University of Oregon
Eugene, OR
butler@cs.uoregon.edu

Carrie Gates
CA Labs
New York, NY
carrie.gates@ca.com

Steven Greenspan
CA Labs
New York, NY
steven.greenspan@ca.com

ABSTRACT

Traditionally, if someone did some act that required forgiveness, there were social norms in place for such forgiveness to happen. Over time, the act is also typically forgotten. And, should the person not be forgiven and the social pressure become too great, he had the option of moving to a new location for a fresh start. Yet with the Internet, these options are no longer available. Worse, activities which traditionally did not even require forgiveness are now impacting lives in unexpected ways, and are never forgotten. There are, however, technical approaches that could be applied to the problem, such as (1) controlling dissemination through new access control models or cryptographic approaches, (2) flooding the web with contrary information, (3) leading users to believe the information applies to someone else, (4) changing the semantics of what was written, and (5) finding a way to take advantage of the inconvenient information. In this paper we discuss the social act of forgiveness, and go into detail on the possible technical approaches to "forgetting" without deleting.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues—*Privacy, Use/abuse of power, ethics, human safety*

Keywords

Digital forgetting; ethics of forgetting; how to digitally forget

1. INTRODUCTION

The Internet can be an unforgiving place. Once data is posted on it, it is immediately indexed, searchable, and preserved for eternity. The question of whether or not this should be the case is a hot button issue that has received much attention in the news, most recently regarding the European Union's proposal of a "right to be forgotten" law [35]. The law, which is part of an expansive set of policies to regulate personal data protections, would essentially allow individuals to withdraw their consent to use personal data they have posted online at any time, thereby putting the onus on those collecting the data, rather than the individual posting the data, to prove why those records should be kept and monitored. Many individuals, lawmakers, and organizations have labeled the deletion of such records tantamount to "rewriting history", "censorship", and promoting "digital death" [18, 19, 33, 46]. As Mayes [33] argues in *The Guardian*, "Being forgotten might sound appealing for some, but making a right out of it degrades the concept of rights. Instead of being something that embodies the relationship between the individual and society, it pretends that relationship doesn't exist".

Most of these equate the concept of "digital forgetting" with "digital deletion" or "rewriting history" [18, 33, 46]. However, this need not be. In the case of rewriting history, the assumption that deletion of public information on the Internet is tantamount to destroying a true record of an event is false. It gives power to search engines to be the ultimate authority on historical accuracy, which is simply not the case. In the case of public records of someone's criminal past, the digital record is not the legal one. Even if information was removed online, the criminal record would not cease to exist. It would simply require individuals to move to offline or private digital repositories (e.g., restricted police records) for tracking down the data they seek. This is no different than in the pre-Internet age.

This ability for information to fade into the archives is what has allowed individuals to let their past mistakes be "forgiven" through being "forgotten". In the past, individuals were able to move and start over in new places without having news of their past indiscretions travel with them. As Etzioni and Bhat [18] also remind us, "The idea that people deserve a second chance is an important American value". In the digital age, however, we are not afforded the luxury of burying past mistakes and starting over. Those mistakes are always one Google search away, disallowing a return to obscurity. Don't those in the digital age have a right to have their "digital sins" be forgiven? If so, is it possible to do so while circumventing the ethical issue of data deletion? In this article, we argue that yes, individuals should have a limited right to have their actions "forgotten" and yes, it is possible to do this not by deleting information but by obfuscating it.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
NSPW'13, September 9–12, 2013, Banff, AB, Canada.
Copyright 2013 ACM 978-1-4503-2582-0/13/09 ...$15.00.
http://dx.doi.org/10.1145/2535813.2535814.

In this paper we examine the social aspects of forgiveness (Section 2), followed by presenting several possible technical approaches to forgetting in Section 3. We then discuss some of the legal and social aspects of applying these technical approaches, such as the notion that we might be "rewriting history" in doing so, along with how these approaches might have negative consequences, followed by conclusions in Section 6. Our contributions include the suggestion of multiple technical approaches to allowing individuals to not be haunted by their online past, including methods to control the dissemination of information (such as originator-controlled access control and cryptographic schemes), deceiving the searcher by hiding the real information amongst decoy information, falsely attributing your actions to another individual with the same name, changing the semantics of the published information, and encouraging the data holder to not disseminate the information.

2. WHY FORGIVE?

Discussions of forgiveness are invariably bound up with issues of privacy and identity, deception and transparency, the public's right to know and the individual's need to be forgiven — to be permitted to re-invent and improve oneself without the burden of past transgressions.

Forgiveness is a fundamental construct in many religions and cultures. In some religions, individuals are commanded not to discuss or gossip about a person's past behaviors; in others, unconditional forgiveness is a sacred virtue. In these instances, the emphasis is on forgiving — acknowledging the wrongdoing, but restoring the individual's place in society. Even when forgiveness is granted, the public expression of forgiveness is often some form of forgetting. The embarrassing or illegal activity is no longer talked about [45], or, as in the case of the tragic shooting in an Amish schoolhouse, the physical reminders are demolished and entirely removed [3, 43]. Although these communities may forgive a transgression, if the transgression is reported on the Web, the rest of the global community may never forget nor forgive the individuals involved.

Early in the evolution of the Internet, the digerati proclaimed a new era in which individuals could explore new identities and interact with one another without the baggage of the past or present. Second Life celebrated the chance to reinvent oneself; people in communities and forums were encouraged to have multiple aliases — to have a business identity, and family identity, and membership in multiple online communities [54]; however, Facebook and other social media sites encourage the blending of business and private lives, and the merged expression of public and private lives [45]. The success of these sites argues for their desirability. The danger is that information from one sphere of life can damage an individual's status in another sphere of life. The leakage from personal to professional, from one community to another, is especially dire in a Web that never forgets nor forgives.

In 2006, a student teacher posted a MySpace photo that showed her wearing a pirate hat and drinking from a plastic cup. The caption read "Drunken Pirate" and may or may not have reflected her true state or what was in the cup from which she was drinking. However, this photo with caption was the basis for denying her a teaching degree. As of 2010, she was employed in another occupation [45]. Her "inappropriate" activity was legal, off-duty conduct; however, the web does not forget, and viewers might not have a sense of humor. What is significant here is that the individual involved was engaged in a legal activity and she was not advised to simply remove the inappropriate photo. Rather, she was denied certification in her chosen profession.

The danger going forward is that newer methods of discovery and correlation will make it easier for businesses, adversaries, and the merely curious to discover unpleasant truths about individuals. We may see in the future an "Equifax" of reputation where employers or banks can go to discover someone's cooperativeness, reliability, activities and affiliations [45]. The scope of tomorrow's search will not be limited to what someone posts on their Facebook site, but also on untagged photos, comments from others, group memberships, credit card purchases, comments made by the person while using another online alias, etc.

There is a tension here between two social needs. One is the social need for transparency and public safety. For example, employers want to know if a prospective hire has a police record, if they engage in drunkenness, if they make negative statements about their employers on blogs, and if others in their profession admire or criticize them. Voters want to know if the politicians they elect act appropriately, and consumers desire information about the businesses, doctors, and products they use. The other is the social need for privacy and to allow individuals a chance to improve themselves, to reinvent themselves, and to interact with others without constant reminders of transgressions from years past. To resolve this tension, we can apply a paraphrase of Etzioni's rubric for disrupting privacy rights [17]:

1. What is the present and future danger to public health and safety if the damaging information is removed or obscured?
2. Can the danger be reasonably countered if the damaging information is not discoverable?
3. If the damaging information should not or cannot be removed or obscured, can we minimize its negative impact on the individual?
4. If the information is removed or obscured or if we minimize its impact, can we treat the undesirable side effects?

In brief, if removing the damaging information benefits the individual and does not harm anyone, then the information should be removed, modified or made difficult to discover. But in a Web that does not forget and encourages connections between data, this is not easily achieved. The solutions that have been offered run the gamut from legal to social to technical: some have argued for laws requiring the removal of all false or slanderous posts when the website owner is notified, but this does not remove true but embarrassing information. Others have argued for laws prohibiting or constraining the use of information about lawful, off-duty conduct in hiring, firing, and promotion decisions [45].

Social solutions tend to be utopian, hoping that transparency will eventually give way to social acceptance of past transgressions that no longer pose a risk. Technical solutions tend to focus on either (1) removing the damaging information after some period of time either by explicitly deleting all known copies of the data or by initially encrypting the information and destroying the encryption key after some duration [21], or (2) making the information difficult to find [44].

For example, `reputation.com` claims to hide negative information by shifting the ranking of online content, making it more likely that links to positive information will be posted in the first several pages of a web search. None of these solutions are perfect and some combination of technical, social and legal advances will be necessary.

It is in society's best interest to care about this issue. If employers only hire those with "clean" slates, they may be surprised to learn that their employees conform but do not push the limits and innovate. If individuals fear reprisals for lawful but embarrassing conduct, they may feel overly constrained and stressed. In a knowledge society that values diversity and creativity, individuals should be allowed to mature, improve and re-invent themselves. The question is how best to achieve that goal.

3. DIGITAL "FORGETTING"

Forgiveness, in this context, encompasses forgetting as well as the more traditional "forgiving". The idea is that by somehow removing the data about the incident that one wishes to be forgiven for, the entire notion of "forgiveness" becomes irrelevant. One cannot find the data about the action[1] that one needs to be forgiven for.

Several countries have passed, or are attempting to pass, "right to be forgotten" laws that mandate providers remove information from the Web and other sources [25, 4, 28]. The basis for what follows is an assumption that such laws will be ineffective for technical, social, political, and legal reasons. We also note that our work covers only data stored on the Web, although it may provide insights useful in other realms. In particular, absent the memory erasing technology described in Philip K. Dick's short story "'Paycheck" [15], memories of actions and events will remain. But the memories can be denigrated as incorrect or misleading in the absence of evidence, and so our work focuses on the evidence.

In the following subsections, we explore several approaches to coping with released information without having the ability to delete it. These approaches range from access control, to cryptographic approaches, to deception, to linguistic approaches.

3.1 Controlling Dissemination

If information about an action cannot be disseminated, then it cannot be found. Therefore, controlling the dissemination of information enables someone to prevent sets of people from seeing that information. This immediately suggests an ORCON model [22].

Originator-Controlled access control gives the "originator" the ability to grant or revoke access to the *information* rather than the containing entity (usually a file). Thus, if Anne "originates" the information about her, and she propagates it to Betty, Caroline, and Dana, Dana can send it on to Elizabeth only with Anne's consent. Furthermore, once Anne decides to revoke access to the data, Betty should not be able to read it any more. Hence our purpose requires a "temporal ORCON" model, in which the originator can revoke access as well as grant or deny it.

The ability to delete (or revoke access to) data may have a number of undesirable consequences. First, any notion of data provenance will be compromised, so the notion of "revocation" must either include changing the relevant provenance, or accepting that the integrity of the provenance must be compromised. It will also affect other data further downstream. For example, if the data is used in a novel, then revoking access to the data requires revoking access to those parts of the novel that use that data. One then must take into account other data derived from the data that is to be suppressed. In essence, we are faced with the problem of revoking access not simply to the data d, but also to transformations of the data $f_1(d), \ldots, f_n(d)$. Determining these outputs from the transformations when the transformations are not known appears to be an insoluble problem; either too much or too little will be suppressed. Furthermore, the situation where the data is being processed raises an additional complication, because the data must be removed from memory; while conceptually this is a detail, in practice it adds complications about tracking the data as it flows through all parts of the system (including peripherals such as video monitors).

Perhaps a different structure of the data would help here. Instead of propagating data, consider propagating links to the data only. This is akin to providing capabilities, which identify the address (location) of the object and the rights that the possessor of the capability has over the object. To simplify revocation, the address is typically indirect, pointing to an entry in a "global object table" that contains the address of the object. Then, to revoke access, the entry in the global object table is altered. This prevents access to the raw data. But tracking the derivatives obtained from that data will require some method of tracking, such as a tamperproof provenance, to avoid the problem noted in the preceding paragraph. How to construct a provenance is beyond the scope of this paper, but we do note that the initial *content* of the provenance is critical [20]. This is an area that needs to be explored in more detail.

Other approaches apply a notion of ephemeralness to the data. That is, after some period of time, the data *and all copies of it* become unavailable. The need to allow data to age is one that businesses are acutely aware of: as information continues to be generated at ever-increasing rates, a hierarchical approach to dealing with less-current data becomes necessary. The practice of *information lifecycle management* (ILM) deals with these concerns and while strategy is not universally codified [12], it is one that many companies use and that databases such as Oracle support. Much like how paper records eventually become consigned to large warehouses in rural areas after their retention period expires, the strategy behind ILM is to move data to less accessible storage as it ages and becomes less immediately relevant, such that old data will eventually be relegated to tape storage and possibly go offline altogether.

Unfortunately, these mechanisms are not available to the regular user, so other methods of allowing them to control data lifetime are necessary. Users are acting as content creators; hence it may be reasonable for them to desire controlling that content through the same types of digital rights management schemes that have been proposed to secure other content. Unfortunately, as we have seen, DRM is often applied in a ham-fisted way that serves to irritate legitimate users rather than providing meaningful content protection, as these protection mechanisms tend to be quickly broken. Perhaps better schemes may be designed or managed when

[1]Technically, one can also need forgiveness for thoughts. As thoughts are not visible to external observers, but actions resulting from those thoughts are, we focus on actions.

DRM mechanisms are decentralized and individual users are the ones managing their rights over the content that they produce. Solving issues such as the "analogue hole", where data can be exfiltrated through other means, will continue to remain an issue, however.

Another way to ensure that data lifetime is respected is for the application or underlying platform itself to enforce data deletion after a period of time. Several such schemes have been proposed [9, 39]. For example, Geambasu *et al.*'s *Vanish* proposal does exactly this, where information can be set to "self-destruct" after a given amount of time. It is based on the concept of shattering a key used to decrypt a piece of data and having the key shards stored in a distributed hash table, where the shards will erode over time due to churn of the nodes and built-in timeouts where DHT nodes purge their data periodically. Without the necessary number of key fragments to reconstruct the decryption key, the data becomes permanently unreadable. Proper implementation of this scheme is extremely important, as weaknesses in the DHT mapping algorithm can allow reconstruction of the key shards. This is the basis behind the *unVanish* attack [62], which allows extraction of key shares from the DHT in an efficient manner before they expire. Services such as SnapChat [1] have been designed to automatically remove material from smartphones a set amount of time after it has been viewed, but even the designers of the app have said that privacy of the data is not its primary purpose, as other means (e.g., the analogue hole) can be used to extract this information, and ways to preserve the picture after it expires are straightforward [16]. Thus, defense mechanisms are complex from an implementation standpoint and circumvention is still a difficult problem to overcome.

It should be noted that there are problems with controlling dissemination. For example, one often does not know what data one wishes to control, and so one has to have mechanisms to enable the control of *all* data. Legal and social problems abound; for example, who owns a criminal record — the convicted criminal, the victims, the police, courts, or some other entity or combination of entities? In a transnational Internet, who decides these questions? Thus, from the practical point of implementation, controlling the dissemination seems impractical.

3.2 Hiding

If controlling dissemination is unsuitable, hiding the action might prove more successful. The idea here is to conceal the action so that the data concerning it is difficult to locate. This can be done in a number of ways.

3.2.1 Deception and Flooding

In litigation, when one party asks for discovery from another party, the second may simply show the first several warehouses of documents (or the electronic equivalent), forcing the first party to spend much time looking through the documents for what they need (and, not coincidentally, making it harder to find that information). This is the idea behind the approach of flooding. To see its effectiveness in computer security, recall that a standard lament of security officers who analyze data from intrusion detection systems is that the amount of low-level data is often overwhelming, and obscures information about serious attacks. For this reason, a number of techniques abstract high-level events from these logs.

The approach that we are proposing is for the subject to release large amounts of similar information that is not correct. The viewer must then pick the correct confidential information from the mass of incorrect information. This requires the production of synthetic, but convincing, data, and then ensuring it is released in a manner convincing to the adversaries who will view this data. This approach has been used successfully in warfare, for example to obscure the destination of the Allied attack on southern Europe in World War II [34, 31]. In that episode, the Allies created a false persona that convinced the German High Command that the attack would come at Sardinia rather than Sicily, which was the obvious point of attack.[2] The point of relevance here is that the information in the documents pointed to disinformation that would aim to convince the Germans that the attack would aim for Sicily. Thus, even if German agents uncovered information about the real target, the deception would lead them to believe the correct information was actually a trick.

Computer decoys are an example of this approach. Stoll's efforts to trace and trap an intruder in the Lawrence Berkeley National Laboratory computer network [51, 50] involved creating a false document that took the attacker several hours to download, distracting him from his search for similar documents. Cheswick [13] used a similar approach to distract an attacker, leading him or her to think that the attacks were successful. Honeypots and honey nets are also examples of these systems, the goal being to draw the attackers onto a controlled platform where they may be monitored and occupied so they do not disrupt the production systems and work.

A good example of a situation where this approach is effective is for the insider attack [11]. Bowen *et al.* [10] suggested using decoy documents to require an attacker to distinguish real information from false information, thereby confusing the attacker; this also allows the defenders to monitor those decoys, and hence the actions of the attackers. Ben Salem and Stolfo [7] characterize the desirable properties of decoys they have found effective. Voris *et al.* [57] discuss how to craft such decoys automatically.

Therefore, to conceal the information, one simply creates and disseminates false information that is very similar to the information to be hidden (or, in this context, one should say "obscured"). The decoy generation tools proposed for the insider problem may help here. Then one trying to discover the true information must determine how to distinguish it from the plethora of false information available. If one can arrange for search engines to feature the false information prominently, then so much the better.

This technique is vaguely reminiscent of phishing, where large numbers of phishing emails are sent out in the hopes that some will find their mark. Often more effective are targeted phishing attacks, where the email is crafted to be credible to a particular person; this is called "spear phishing". A similar approach for deception is appropriate here: aim at a specific part of the information, specifically whom it is about.

3.2.2 False Attribution

A variant of this technique is to attribute the act to someone else. This leaves the information intact, but changes the

[2]Indeed, Churchill said that "anybody but a damn' fool would *know* it is Sicily" [34, p. 24]

target of the information. Ethically, this approach is highly questionable; but a discussion of the conditions under which such diversions are unethical is outside the scope of this paper.

Consider a report that "Matt Bishop wrote a paper on cyberwarfare." The author is ambiguous; is it the computer scientist Matt Bishop of the University of California at Davis, the political scientist Matt Bishop of the University of Sheffield, or one of the myriad of other Matt Bishops in either field? Similarly, if one reads in a blog that "Carrie Gates likes electronic music", is that the computer scientist Carrie Gates or the VJ Carrie Gates? In all cases, the confusion is accidental because two people have the same name. But such confusion could also be created deliberately—especially in these days of on-line identities that can be bogus.

Identity management schemes enable the disambiguation of identities, and so would seem to solve this problem. In theory, they could; but several practical considerations arise. First, there is no universal naming scheme that everyone subscribes to, and given the experiences with certificate management systems, such a scheme is unlikely ever to be adopted. Thus, while unique naming may be effective for members of an identity management system, it will not provide adequate identification for those not in the system. Second, when one joins such an identity management system, what steps are taken to ensure that the identification is correct *initially*? History is replete with examples of impersonations, and the Web is a fertile source for identity theft, which enables such impersonation. Third, people move on, and often do not maintain their membership in identity systems so that the information is up to date. If identity is bound to a role, for example, such actions could result in mis-identifications.

In essence, this is a problem of attribution. The identity of the entity to which the data refers is bound to the data in some way (possibly by being part of the data, possibly by being external to but bound to the data). So the idea is to change the *referent* of the bound entity, and not the binding. This requires obscuring or changing the interpretation of the identification of the entity.

This also suggests a more general approach: reinterpret the data itself.

3.2.3 Change the Semantics

For the third approach, imagine again the situation of the teacher who was denied promotion partially due to the "drunken pirate" picture surfacing of her on the Internet [45]. Although there was no proof that what the teacher was drinking was actually alcoholic or not, the potential "inappropriateness" of her behavior in the photograph due to the caption was enough to render her "guilty". Imagine again the same picture, instead with a caption reading "Emily having fun at Kevin's birthday party" or "Look how brave Emily is for trying Kevin's homemade Jamba Juice". Would the teacher likely have received such harsh judgment? Probably not.

Contextual ambiguity and impression formation has long been the bane and boon of computer-mediated communication, and has long been the subject of research in Social Psychology, Applied Linguistics and Communication Studies [8, 32, 42, 48]. Much research has investigated the breadth/depth of impressions online [24, 59, 60], extralinguistic factors influencing impressions (e.g. skepticism, paralanguage, sociability) [56], and linguistic factors influencing impressions [27, 37, 52]. And despite the lack of nonverbal and social context cues available in computer-mediated communication, research has shown abundant evidence of the development of complex, interpersonal relationships online, even in the absence of such cues [38, 58, 61]. Although the research has been decidedly agnostic regarding whether this contextual ambiguity is a good or bad thing, few if any have looked at the possibility of manipulating linguistic cues to increase ambiguity, and the social reasons for doing so. However, by changing the possibility of what a picture or situation could plausibly represent, we could introduce enough uncertainty about what was really occurring as to protect oneself from a potentially damaging attack using that picture/situation in the future.

One way that we might be able to do this systematically in the future is to harness the increasing number of corpus-based semantic models (CSMs) being designed to identify semantic relationships between words [5, 30, 29]. CSMs are designed to scan large bodies of naturally occurring texts and extract information about both the contextual meaning of particular words within those texts as well as their semantic similarity to other words within those texts. However, as Baroni *et al.* [5] point out, "CSMs might be very good at finding out that two concepts are similar, but they tell us little about the internal structure of concepts and, hence, why or how they are similar" [5, p. 223]. With the design of CSMs like Strudel [5], researchers have begun not only to identify which words have relationships with one another but also have automatized what the semantic relationships between those words are. For example, now we cannot only identify that dogs are related to cats and the other words/phrases that co-occur with these terms in different contextual situations, but that the salient "properties" of dogs are that they fit into the category of animals and that typical behavior is that they bark. This is a departure from other CSMs designed to extract contextual features of words from texts [2, 14, 40, 41] because the potential identifying features do not have to manually entered and customized in advance.

With photographs, the range of potential meanings to explain the story of the photo is vast. By identifying the contextual and linguistic features tagged in the original photo, it could help us identify other plausible contextual situations to explain the same photo. While Baroni *et al.* [5] admit that Strudel and other models cannot yet handle polysemous words[3], these types of algorithms take us one step closer to automatically identifying linguistic phrases and contexts that could be plausibly used to retag "offending" or "dangerous" photos or comments surfacing online for which we want to provide alternative interpretations.

3.3 Inconveniencing the Interpreter

An approach that has proved effective in other environments is to accept that the information is now known and, in effect, make it prominent enough in some way to gain some advantage from its leaking. This does not hide the confidential information, but it can illuminate previously unknown adversaries. For example, one then can set traps to detect an adversary exploiting that information, perhaps by adding markers or bogus information that will reveal an attacker trying to exploit it. Stoll's work, mentioned earlier, the de-

[3]Polysemous words are words that have multiple similar meanings. [55]

coy work in insider detection, and honeypots and honeynets are all examples of this.

An alternative approach is to tie it to other information so that the *combined* data produces an unbelievable interpretation, or will inhibit others from promulgating it. Tip O'Neill recounts such an incident in his autobiography [36], when reporters for the Boston Globe tried to obtain a list of donors, which O'Neill refused to provide, in order to protect the donors from being approached by others for money. The resulting story suggested that O'Neill had something to hide. O'Neill immediately sent the list of donors, and amounts, to the publisher of the paper, saying he was welcome to print all of the names — but only if he printed *all* the names. The publisher of the paper was one of the donors in the list. The story was dropped.

This technique clearly will not always work. The risk of the adversary distinguishing the false information from the accurate information, or of deciding the embarrassment of exposing the ancillary information is worth exposing the private information, may make this tactic untenable. But when it does work, this tactic is remarkably satisfying.

3.4 Metrics for Forgetting

Quantifying the effect to which information can be effectively forgiven and forgotten is a very difficult problem for which metrics are not readily derived. One way in which the security community has considered the quantification of costs has been through the concept of "work factor," an attempt to quantify the cost of breaking the security of a system. While such an approach can be related to computational complexity in the cryptography community, it is less obvious how to apply this to other areas of security. Saltzer and Schroeder [47] considered the principle of work factor and noted that many protection mechanisms cannot be directly calculated with it. Moreover, estimating these calculated values can be extremely difficult.

Additionally, many of the measures that we are considering may not be directly quantifiable using a single metric, or its value may be less apparent. For example, one could consider an information-theoretic approach to quantifying data lost or "forgotten," but such an approach would not be able to properly characterize a measure such as changing the semantics of the information presented. Information in this case is not lost, but rather re-contextualized, and reducing this to a single metric compatible with approaches such as flooding may not be possible. As a result, the solution to assigning a metric may require a vector rather than scalar quantity, with the metric adapted to a particular scheme. Even subjective measurements of work produce different results when evaluated by different workload scales and are dependent on context for an appropriate measurement [26].

It may be possible to attempt to quantify the work involved in performing any particular technique such as the effort involved in designing a flooding scheme or developing a new semantic model if such tasks are considered activities that could be profiled with cognitive metrics [23], but determining the outcome of these approaches and assigning quantifiable metrics to the efficacy of the results remains an open research problem.

4. GENERAL DISCUSSION

The notion of being able to "forgive and forget" individuals on the Internet opens up a number of non-technical issues in addition to the development of possible technical approaches. And these issues may prove to be more difficult than technical solutions.

The first issue is an inherent assumption about who owns the information. If you want to be forgiven for something, and for the world to forget that it happened rather than having perpetual reminders, then it implies that you own this information about yourself, and should therefore be able to take steps to eliminate it. But is this actually the case? For example, the person who stores the information — and therefore has access to it — can be said to "own" the information, given that they have the ability to delete or modify it. But what if this information is about another person? Beyond this, who controls the interpretation of the information? The same piece of factual information can be written such that one has sympathy and understanding for the individual's plight, or that one vilifies the individual.

Further complicating matters is a legal notion of forgetfulness, which varies by country. For example, under German privacy law, a person has the right for their name and likeness to be protected. There is a recent example of a convicted murderer suing Wikipedia to have his name removed as having been the person who murdered actor Walter Sedlmayr, citing German privacy laws [28]:

> "Our client has served 15 years of his life sentence for murdering Mr. Sedlmayr in 1990. He has been released on parole [sic] in August 2007. His rehabilitation and his future life outside the prison system is severely impacted by your unwillingness to anonymize any articles dealing with the murder of Mr. Sedlmayr with regard to our client's involvement," according to the Oct. 27 cease-and-desist letter, which demands legal fees and compensation for "emotional suffering."

Should the Internet have functionality that allows/forces it to adhere to the differing (censorship) laws of various countries? Should there be overlays of "national internets" where local filters state what information can be downloaded? Should a legal notion of forgive and forget be enforced nationally, if not internationally? And, at what point do such laws result in rewriting history? As noted by the Electronic Frontier Foundation [28]:

> At stake is the integrity of history itself. If all publications have to abide by the censorship laws of any and every jurisdiction just because they are accessible over the global Internet, then we will not be able to believe what we read, whether about Falun Gong (censored by China), the Thai king (censored under lese majeste) or German murders.

While rewriting history is not a new concept (see, for example, [63], which describes the "necessity" of rewriting history books in Russia after the fall of communism and the impact of historical perspectives on Russian identity, and [49], which examines efforts to re-evaluate the past to understand the "truth" of events and the approaches used to do so) there seems to be a movement that expects that what is found on the Internet is fact, and that such facts are not presented in the context of any cultural bias, nor as a manipulation for political or financial gain. For example, studies have found that undergraduate students with low evaluativism are less

likely to take into account author perspectives when reading online material [6].

In fact, our views on information, what is "fact", how (or if) information should be presented, and even legal rulings, all reflect our cultural norms. Thus there is also a social — and related ethical — component to the notion of "forgive and forget". How can a system be designed that takes into account varying social expectations? Or . . . should we avoid designing such a system at all and allow new expectations and norms to develop?

5. NEGATIVE SALIENCE

As with all technologies, these approaches can also be used to harm individuals in addition to providing them with recourse against negative depictions and information. We can look at the above approaches and divide them essentially into two categories — reprioritization of facts or the insertion of fiction. The positive results from the reprioritization of facts can be seen from already existing web sites such as `reputation.com` [44], which promotes positive information and demotes (or forgets) negative information. In terms of fiction, it is possible to create fake friends, fake references, etc. It is also possible to recast negative events (or events *perceived* as negative by some) into a more positive frame. For example, the drunken pirate photo referenced in Section 2 could be described as being part of a "fake" college play.

With both of these categories, though, it is also possible to have negative salience. For example, it is easy to imagine an anti-reputation service that would increase the likelihood of negative facts surfacing, regardless of the accuracy of those facts. One example of this already occurring is "revenge porn" sites, where angry exes post naked pictures of their now ex-girlfriends or boyfriends. In this case, negative "facts" (pictures) are used. In other cases, fiction can be used, such as by spurned lovers seeking retribution. An example case of this is a Canadian teacher who is unable to find work due to cyber-stalking and cyber-bullying from an ex-girlfriend who has posted untrue comments about the teacher on various web sites, while the teacher is unable to have the comments removed [53]. Interestingly one of the commenters (Digital Culture) on this article suggests equally flooding the Internet with positive content.

6. CONCLUSIONS

Individuals within society have traditionally enjoyed the ability to both forgive and be forgiven for social transgressions. Over time, such transgressions were generally forgotten, or at least not the subject of further discussion. In those cases where the social sin was great enough to not be forgiven, a person could always relocate and start anew, without having his past follow him to his new setting. So, while not forgiven, the act was essentially forgotten.

However, in today's world, where with one upload an image or text is indexed, cached, copied, made easily searchable, one can never escape their digitized past. Thus, even for an event where one would traditionally have been easily forgiven, there are constant online reminders of the transgression, and new acquaintances, potential employers, recent co-workers, are all able to easily find this information with a simple search on the individual. Worse, previously private events are now easily found in the public domain (such as pictures from parties), and the individual can be judged anew by people who do not even know the context.

There are technical approaches to minimizing undesirable digital information, such as using ORCON, flooding or obscuring the real information with new or different information, changing the semantics of the public information, deceiving the user with false information, and aging data using cryptographic protocols. However, there are limitations to all of these approaches; more specifically there are no solutions that can guarantee obscurity for the user. That is, while technology can make life easier, you cannot put the proverbial genie back in the bottle once that information has been made publicly available.

An alternative method is not to change the data or history, and not to obscure it. Rather, we can break the binding between *most of the data* associated with an individual and the individual identified. A new identity would be created for the person essentially by (legally) changing their name. As this does not break the tie with their social security number, they could still have significant parts of their history move with them (e.g., educational accomplishments); however, this can be an expensive procedure and would likely be imperfect (although we note that women getting married have been doing something similar for many decades!). In an age with sophisticated image search capabilities, more complete identity change might also require physical alternations. Nonetheless, it is the closest we can come to "moving to a new village", which in its day was also expensive.

The techniques discussed in this paper essentially make uncovering the truth harder and more expensive. However, deployment of these and other techniques depend upon individuals having access to the appropriate economic, political, and technological resources. Individuals fortunate enough to have this access, and the ability to apply these resources, will be "forgiven", but those less fortunate will not be "forgiven". For example, even now not everyone can afford to hire consultants to protect their reputation, and many are not even aware that these services exist. Legal remedies, such as the German privacy laws, attempt to provide rights to all citizens, but for reasons discussed above such remedies are imperfect, create side effects, and access to legal counsel needed to apply the legal remedies can be costly.

Even if a perfect technical solution were available, society needs to determine — on a global scale — how forgiveness needs to be provided, and what forgetting means — is it the deletion of data, or some other mechanism (such as being available only in public records that are *not* kept online)? And society needs to determine if, ultimately, not having information available online in a searchable format is the equivalent of rewriting history.

Acknowledgments: Matt Bishop acknowledges the support of awards CCF-0905503, CNS-1049738, CNS-1219993, and CNS-1258577 from the National Science Foundation to the University of California at Davis for support. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation or of the University of California at Davis. Kevin Butler acknowledges the support of NSF awards CNS-1254198 and CNS-1118046.

7. REFERENCES

[1] Snapchat. http://www.snapchat.com, 2013.

[2] A. Almuhareb and M. Poesio. Attribute-based and value-based clustering: An evaluation. In *Proceedings of EMNLP*, pages 158–165, 2004.

[3] M. Ambrose, N. Friess, and J. Van Matre. *Seeking Digitial Redemption: The Future of Forgiveness in the Internet Age*. 2012. December 13.

[4] Nate Anderson. Spain Asks: If Google Search Results Make Your Business Look Bad, Can You Sue? Ars Technica News Report, February 2012.

[5] M. Baroni, B. Murphy, E. Barbu, and M. Poesio. Strudel: A corpus-based semantic model based on properties and types. *Cognitive Science*, 34:224–254, 2010.

[6] Sarit Barzilai and Yoram Eshet-Alkalai. The role of epistemic thinking in comprehension of multiple online source perspectives. In *Proceedings of the Chais conference on instructional technologies research 2013: Learning in the technological era*, pages 11–17, 2013.

[7] Malek Ben Salem and Salvatore J. Stolfo. Decoy document deployment for effective masquerade attack detection. In Thorsten Holz and Herbert Bos, editors, *Proceedings of the 8th International Conference on the Detection of Intrusions and Malware, and Vulnerability Assessment*, volume 6739 of *Lecture Notes in Computer Science*, Berlin, Germany, August 2011. Springer-Verlag.

[8] C.R. Berger and R.J. Calabrese. Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication. *Human Communication Research*, 1:99–112, 1975.

[9] Nikita Borisov, Ian Goldberg, and Eric Brewer. Off-the-record communication, or, why not to use pgp. In *Proceedings of the 2004 ACM Workshop on Privacy in the Electonic Society*, pages 77–84, 2004.

[10] Brian M. Bowen, Shlomo Hershkop, Angelos D. Keromytis, and Salvatore J. Stolfo. Baiting inside attackers using decoy documents. In Yan Chen, Tassos D. Dimitriou, and Jianying Zhou, editors, *Proceedings of the 5th International ICST Conference on Security and Privacy in Communication Networks*, volume 19 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 51–70, Berlin, Germany, September 2009. Springer.

[11] R. Brackney and R. Anderson. Understanding the Insider Threat: Proceedings of a March 2004 Workshop. Technical report, RAND Corporation, Santa Monica, CA, March 2004.

[12] Ying Chen. Information Valuation for Information Lifecycle Management. In *ICAC 2005: Proceedings of Second IEEE International Conference on Autonomic Computing*, pages 135–146, 2005.

[13] Bill Cheswick. An Evening with Berferd, in Which a Cracker Is Lured, Endured, and Studied. In *Proceedings of the Winter 1992 USENIX Conference*, pages 163–174, Berkeley, CA, USA, 1992. USENIX Association.

[14] P. Cimiano and J. Wenderoth. Automatically learning qualia structures from the web. In *Proceedings of the ACL/SIGLEX Workshop on Deep Lexical Acquisition*, pages 28–37, 2005.

[15] Philip K. Dick. *Paycheck and Other Classic Stories*. Citadel, New York, NY, USA, 2003.

[16] Zach Epstein. Your Snapchats Aren't Safe: How to Secretly Save Videos from Snapchat or Facebook's 'Poke'. http://bgr.com/2013/01/01/how-to-save-snapchat-video-secretly-278227/, January 2013.

[17] A. Etzioni. *The Limits of Privacy*. New York: Basic, 1999.

[18] A. Etzioni and R. Bhat. Second chances, social forgiveness, and the internet, 2009. http://theamericanscholar.org/second-chances-social-forgiveness-and-the-internet/#.UWOSWxnB1xM.

[19] P. Fleischer. http://peterfleischer.blogspot.com/2011/03/foggy-thinking-about-right-to-oblivion.html, 2011. http://peterfleischer.blogspot.com/2011/03/foggy-thinking-about-right-to-oblivion.html.

[20] Carrie Gates and Matt Bishop. One of these records is not like the others. In *Proceedings of the 3rd USENIX Workshop on the Theory and Practice of Provenance*, Berkeley, CA, USA, June 2011. USENIX Association.

[21] Roxana Geambasu, Tadayoshi Kohno, Amit A. Levy, and Henry M. Levy. Vanish: Increasing data privacy with self-destructing data. In *Proceedings of the 18th USENIX Security Symposium*, pages 299–315, 2009.

[22] R. Graubert. On the need for a third form of access control. In *Proceedings of the 12th National Computer Security Conference*, pages 296–304, Oct. 1989.

[23] Wayne D Gray, Michael J Schoelles, and Chris Sims. Cognitive metrics profiling. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 49, pages 1144–1148. SAGE Publications, 2005.

[24] J.T. Hancock and P.J. Dunham. Impression formation in computer-mediated communication revisited: An analysis of the breadth and intensity of impressions. *Communication Research*, 28(3):325–347, 2001.

[25] Leslie Harris. Escaping your online mistakes: Should there be a law? ABC News Report, July 2011.

[26] Susan G Hill, Helene P Iavecchia, James C Byers, Alvah C Bittner, Allen L Zaklade, and Richard E Christ. Comparison of four subjective workload rating scales. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 34(4):429–439, 1992.

[27] D. Jacobson. Impression formation in cyberspace: Online expectations and offline experiences in text-based virtual communities. *Journal of Computer-Mediated Communication*, 5(1), 1999. http://www.ascusc.jcmc/vol5/issue1/jacobson.html.

[28] David Kravets. Convicted murderer sues Wikipedia, demands removal of his name. http://www.wired.com/threatlevel/2009/11/wikipedia_murder/, 2009. November 11.

[29] T.K. Landauer and S.T. Dumais. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104(2):211–240, 1997.

[30] K. Lund and C. Burgess. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behaviour Research Methods*, 28:203–208, 1996.

[31] Ben Macintyre. *Operation Mincemeat*. Harmony Books, New York, NY, USA, 2010.

[32] G. Mantovani. Internet haze: Why new artifacts can enhance situation ambiguity. *Culture & Psychology*, 8(3):307–326, 2002.

[33] T. Mayes. We have no right to be forgotten online: The European union is planning to enshrine in law the right to be forgotten on the internet, but we can't live outside society, 2011. `http://www.guardian.co.uk/commentisfree/libertycentral/2011/mar/18/forgotten-online-european-union-law-internet`.

[34] Ewen Montagu. *The Man Who Never Was*. J. B. Lippincott Company, New York, NY, USA, 1953.

[35] No Author Provided. Commission proposes a comprehensive reform of data protection rules to increase users' control of their data and to cut costs for businesses, 2012. `http://europa.eu/rapid/press-release_IP-12-46_en.htm?locale=en#PR_metaPressRelease_bottom`.

[36] Thomas P. O'Neill. *Man of the House*. Random House, New York, NY, USA, 1987.

[37] P. O'Sullivan, S. Hunt, and L. Lippert. Mediated immediacy: A language of affiliation in a technological age. *Journal of Language and Social Psychology*, 23(4):464–490, 2004.

[38] M.R. Parks and K. Floyd. Making friends in cyberspace. *Journal of Communication*, 46:80–97, 1996.

[39] Radia Perlman. The ephemerizer: Making data disappear. *Journal of Information System Security*, 1(1):51–68, 2005.

[40] M. Poesio and A. Almuhareb. Identifying concept attributes using a classifier. In *Proceedings of the ACL Workshop on Deep Lexical Semantics*, pages 18–27, 2005.

[41] M. Poesio and A. Almuhareb. Extracting concept descriptions from the web: The importance of attributes and values. In P. Buitelaar and P. Cimiano, editors, *Bridging the gap between text and knowledge*, pages 29–44. IOS Press, Amsterdam, 2008.

[42] T. Postmes, R. Spears, and M. Lea. Breaching or building social boundaries? Side-effects of computer-mediated communication. *Communication Research*, 25(6):689–715, 1998.

[43] Raffaele. Crew razes Amish schoolhouse, 2006. October 13.

[44] Reputation.com. `http://www.reputation.com\myreputation`, 2013.

[45] J. Rosen. The web means the end of forgetting, 2010. July 21.

[46] J. Rosen. The right to be forgotten. *Stanford Law Review*, 64:88–92, 2012.

[47] Jerome H. Saltzer and Michael D. Schroeder. The protection of information in computer systems. *Proceedings of the IEEE*, 63(9):1278–1308, September 1975.

[48] R. Spears and M. Lea. Social influence and the influence of 'social' in computer-mediated communication. In M. Lea, editor, *Contexts of computer-mediated communication*, pages 30–65. Harvester Wheatsheaf, London, 1992.

[49] Gabrielle M. Spiegel, editor. *Practicing History: New Directions in Historical Writing after the Linguistic Turn*. Routledge, 2005.

[50] Cliff Stoll. *The Cuckoo's Egg: Tracking a Spy Through the Maze of Computer Espionage*. Pocket Books, New York, NY, USA, Sep. 2005.

[51] Clifford Stoll. Stalking the wily hacker. *Communications of the ACM*, 31(5):484–497, May 1988.

[52] L.C. Tidwell and J.B. Walther. Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: Getting to know one another a bit at a time. *Human Communication Research*, 28(3):317–348, 2002.

[53] Kathy Tomlinson. Escaping your online mistakes: Should there be a law? CBC News Report: British Columbia, May 2013. `http://www.cbc.ca/news/canada/british-columbia/teacher-powerless-to-stop-ex-girlfriend-s-cyberstalking-1.1314610`.

[54] S. Turkle. *Life on the Screen*. Simon & Shuster, 1997.

[55] UsingEnglish.com. `http://www.usingenglish.com/glossary/polysemy.html`, 2013.

[56] S. Utz. Social information processing in muds: The development of friendships in virtual worlds. *Journal of Online Behavior*, 1(1), 2000.

[57] Jonathan Voris, Nathaniel Boggs, and Salvatore J. Stolfo. Lost in translation: Improving decoy documents via automated translation. In *Proceedings of the 2012 IEEE Symposium on Security and Privacy Workshops*, pages 129–133, May 2012.

[58] J.B. Walther. Interpersonal effects in computer-mediated interaction: a relational perspective. *Communication Research*, 19(1):52–90, 1992.

[59] J.B. Walther. Impression development in computer-mediated interaction. *Western Journal of Communication*, 57(4):381–398, 1993.

[60] J.B. Walther. Anticipated ongoing versus channel effects on relational communication in computer-mediated interaction. *Human Communication Research*, 20(4):473–501, 1994.

[61] J.B. Walther. Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, 23:1–43, 1996.

[62] Scott Wolchok, Owen S. Hofmann, Nadia Heninger, Edward W. Felten, J. Alex Halderman, Christopher J. Rossbach, Brent Waters, and Emmett Witchel. Defeating Vanish with low-cost sybil attacks against large dhts. In *Proceedings of the 17th Annual Network and Distributed System Security Symposium*, 2010.

[63] Joseph Zajda and Rea Zajda. The politics of rewriting history: New history textbooks and curriculum materials in Russia. *International Review of Education*, 49(3–4):363–384, July 2003.