Digital immolation: new directions for online protest

(pre-proceedings version)

Joseph Bonneau
University of Cambridge Computer Laboratory
jcb82@cl.cam.ac.uk

May 2, 2010

Abstract

The current literature and experience of online activism assumes two basic uses of the Internet for social movements: straightforward extensions of offline organising and fund-raising using online media to improve efficiency and reach, or "hacktivism" using technical knowledge to illegally deface or disrupt access to online resources. We propose a third model which is non-violent yet proves commitment to a cause by enabling a group of activists to temporarily or permanently sacrifice valuable online identities such as email accounts, social networking profiles, or gaming avatars. We describe a basic cryptographic framework for enabling such a protest, which provides an additional property of binding solidarity which is not normally possible offline.

1 Introduction

Throughout history, new forms of technology have been adapted to facilitate and advance social movements and protests [9]. The potential of the Internet to revolutionise social movements by providing low-cost global communication has been analysed by scholars in a wide range of disciplines [7]. In their influential 1996 essay "Electronic Civil Disobedience" the Critical Art Ensemble argued that the Internet enabled powerful organisations to avoid having a conspicuous physical location such as a castle, capitol building, or corporate headquarters, and thus effective protest would eventually have to be carried out by electronic means to be effective [6].

Dorothy Denning's 2001 essay introduced three broad categories of online social movements: activism, hacktivism, and cyberterrorism [3]. Online activism describes using computer-mediated communication to aid in peaceful, legal social movements. Activism techniques include fund-raising, gathering and disseminating information, and signing online petitions. These tactics utilise the decentralised nature of the web to facilitate grass-roots movements whose dispersed supporters may be difficult to organise in the offline world. The low cost of Internet communication can also harm online activism, as individual protests struggle to compete for attention online [9].

The recent growth of online social networks and virtual worlds has opened up new means of activism and non-violent protest. For example, millions of users declare their devotion to various causes on social networks by joining groups or adding icons to their profiles [14], while virtual world participants conduct online pickets outside unpopular organisations' virtual property [1]. Such protests are different from those typically analysed in past academic literature in that they use the existence of online identities to show support for their causes. Social networks have also been widely hailed for their role in organising offline protest. In this function they serve as a highly efficient communication medium, enabling qualitatively different types of movements such as flash mobs which are organised on very short notice.

In contrast, hacktivism entails directly harming an opposition's online resources. A variety of tactics exist [2, 15] including defacing or disabling web sites, blocking access through denial of service attacks, manipulating search engine results, or harassing individuals through email floods. The ethics and legality of such tactics are complicated. Adherents claim that such means are ethical because, as they are not intended to cause any physical violence or harm, they are a modern instantiation of the long and widely-accepted tradition of non-violent civil disobedience [12, 17]. Opponents argue that such tactics violate property rights and are the online manifestations of violent protests [8].

In either case, the results of these protests have been mixed. The earliest large-scale movement, a series of denial of service attacks in 1998 by a group called Electronic Disturbance Theatre supporting the Zapatista movement in Mexico, brought large-scale media attention to its cause [10], but subsequent movements have not always received favorable news coverage. A central problem remains that such tactics demonstrate technological expertise but do little to demonstrate the extent of a cause's support.

Cyberterrorism is distinguished from hacktivism in that it attempts to initiate real-world harm by using the Internet to destroy physical infrastructure. An act of cyberterrorism might be using a remote exploit in a control server to destroy a power station, or cause an airplane crash by disrupting air traffic control systems. Despite the large potential loss and the stated desire of known terrorist groups to conduct such attacks, there have been no successful incidents and the threat remains largely speculative [4].

1.1 Common problems

Both online activism and hacktivism have drawbacks which are increasingly limiting their effectiveness. Online activism struggles to gain attention given the enormous quantity of information available online and the ability of Internet users to filter out information from groups they disagree with [16]. Hacktivism suffers both from increasingly effective technical defences and the possible loss of public support due to its adversarial nature.

In the abstract, a protest aims to demonstrate both *commitment* and *morality*.

1.1.1 Commitment

Commitment requires demonstrating both the existence of a large number of supporters and a serious level of support from each user. For example, a street rally shows that a large number of adherents are willing to expend free time marching in support of a cause. Viewed as a commitment signal, the importance of numerical turnout makes sense. Indeed, protest organisers routinely claim higher turnout than official estimates.

Numbers alone don't prove commitment, there must be a real opportunity cost for participating members. Benefit concerts may attract large numbers of attendees, but still not demonstrate commitment if attending the concert is desirable independent of the specific cause. In contrast, street rallies in inclement weather or with a high risk of arrest demonstrate higher commitment.

Online protests struggle with commitment due to the extremely low cost of signing an online petition or joining a social networking group.[1] Detecting the commitment of protesters is related to the computer security challenge of preventing spam. The low cost of email allows individuals to overuse attention resources for personal gain, requiring automated filtering techniques. Similarly, the low cost of organising an online protest means that many movements can gain a token level of support online. People have limited attention resources for social movements as well, often relying on the mainstream media as a filter to ignore online protests unless they can demonstrate sufficient support.

Ideally, we desire a mechanism for online protesters to send an unforgeable signal of dedication to a cause. For email spam, proof-of-work schemes allow senders to show that a message is genuine by performing costly work before sending. The idea has not caught on in practice and economic analysis has suggested that spammers can expend more effort to send spam messages than ordinary users can for real messages [11]. We avoid computational work as a signal for dedication to a protest because some individuals have much greater computational resources than others. Protests traditionally embrace an ethic of equality, enabling all participants to contribute equally to the signal of commitment.

The innate desire to show commitment can be seen in the creative attempts of some social networking movements to strengthen their signal. An informal protest movement over perceived discrimination against Barack Obama's 2008 presidential campaign saw thousands of users changing their middle name to "Hussein" in solidarity. Anger over NBC's 2010 firing of talk show host Conan O'Brien similarly saw thousands of users changing their profile picture to an image of O'Brien. Such methods show slightly higher commitment, as they diminish the core purpose of the user's profile to showcase information about themselves. Furthermore, they can only be used for one cause at a time, in contrast to group memberships of which users can hold dozens. Still, the level of commitment shown is relatively low and more effective protests would allow a signalling from a much larger spectrum of commitment.

1.1.2 Morality

Many protest movements aim to maintain a moral high ground by maintaining an ethical code which they consider superior to their opposition's. Certainly, protests which aim to succeed through media coverage rather than by inflicting direct harm frequently attempt to claim a principle of "non-violence" to demonstrate that they are being oppressed by some opposition group. There is a vigorous scholarly debate over the ethics of some forms of hacktivism, and adherents frequently attempt to follow a "hacktivist ethic" to avoid such criticism [17]. The FloodNet software introduced in the Zapatista protests and since used in many others, for example, requires many users to download and install software to generate web requests, which they claim makes their protest akin to a virtual "sit-in"

[1] Hacktivism can be even worse at demonstrating commitment, as a single competent hacker may be behind an entire denial of service attack if he controls a large botnet.

with each user consuming only a reasonable amount of resources. Regardless of technical specifics or definition of violence, however, hacktivism remains combative in nature which can prevent its adherents from gaining public sympathy as an oppressed party.

1.2 A new direction

We draw our inspiration from the history of non-violent social protests which were effective because participants faced risks of serious personal harm. This harm could either be from an opposing power structure, in the case of street marches being violently suppressed, or from the participants themselves. Hunger strikes remain a powerful symbol of dedication in that participants provide a real threat to starve themselves to death in support of their cause. An even more extreme example is self-immolation, such as the public suicide of the Buddhist monk Thích Quảng Đức in protest of the Vietnam war.

We propose that online protests can prove strong commitment in a morally-acceptable way by enabling users to temporarily or permanently sacrifice their online identities. We propose a basic cryptographic framework for such a protest and described the key divergences from its offline inspiration.

2 Protocol

2.1 Assumptions

We assume that all participants in the protocol hold an online identity which they are willing to suspend as a part of the protest. This identity may be a social networking profile, and email account, an avatar in a virtual world, or a reputable account in an online market. Each user u_i holds some secret password x_i which controls access to their online identity. Each user also has a separate public/private key-pair $(k_i^{\text{pub}}, k_i^{\text{priv}})$, with all public keys known by all users.

We further assume that a password oracle $\mathcal{O}$ exists which maintains a password table $T : \mathbb{Z} \rightarrow \{0,1\}^*$ mapping each integer user ID i to an arbitrary password. $\mathcal{O}$ allows any user u_i to change their password after completing a zero-knowledge proof of knowledge of x_i (for example, a challenge-response protocol) and sending a new password x_i' to $\mathcal{O}$. It is critical that $\mathcal{O}$ is indifferent to the protest and will not actively oppose it, as we will discuss in Section 3.1.

2.2 Initiation

We require a trusted protest initiator P. The initator does not need to be the actual leader of the group, but needs to be trusted by all protesters to faithfully complete two critical initialisation tasks. We envision that it is best if P is in fact indifferent to the cause, serving the role of an elder statesmen or neutral arbiter.

In signing up for the protest, all participants send their secret knowledge x_i to P along with k_i^{pub}. After all n participants have signed up, P then completes the password update process with $\mathcal{O}$, changing each participant's password to a random value x_i' and then deleting the original value x_i. At this point, all participants have lost the ability to access their accounts, as only P knows the current passwords.

P generates a master key pair $(k_*^{\text{pub}}, k_*^{\text{priv}})$ and n shares $s_1 \ldots s_n$ of k_*^{priv} using a threshold decryption scheme [5]. P then creates a symmetric escrow key k_{e}, and publishes:

$$\left\{ \mathrm{E}_{k_{\text{e}}} \left(\mathrm{E}_{k_i^{\text{pub}}}(x_i') \right) \middle| 1 \leq i \leq n \right\}, \quad \mathrm{E}_{k_*^{\text{pub}}}(k_{\text{e}})$$

This information includes a doubly-encrypted version of each user's new password, under both the escrow key and each users' individual public key, and an encryption of the escrow key to the distributed master key. Upon publishing this information, P has completed its organisation of the protest and can destroy all of its secret knowledge, in particular the new passwords x_i' and the master private key k_*^{priv}.

2.3 Completion

Participants can agree to end the protest by decrypting the escrow key (k_{e}), enabling all users to securely recover their new passwords, after their demands have been satisfactorily met or they have received sufficient media attention. The threshold scheme paramaters can be chosen to require some desired threshold m of n users to agree to end the protest.

It is important to recognise technical differences between threshold decryption and voting. If participants may hold multiple votes and change their votes between rounds, it is necessary to use a homomorphic threshold scheme [5], which enables users to update their shares after each vote so that information cannot be combined from multiple votes.

This protest protocol differs from most of its offline inspirations in that the solidarity of participants is binding. Participants cannot back out of the protest without a quorum willing to end it, unlike traditional protests which rely on maintaining a strong esprit de corps to prevent defection. One binding example from the real world is protesters chaining themselves around a tree, though this can usually be broken by any two participants.

2.4 Variants

2.4.1 Suicide option

The basic scheme only allows voters to end the protest by restoring access to all. It may go on forever if m voters never agree to end it. In a real-world hunger strike, there is a biological deadline limiting the maximum length of the protest. This is difficult to translate into an online protest, but we might assume that the password oracle $\mathcal{O}$ will enable the setting of an account destruction key y_i for each account. In this case, P will generate two escrow keys, k_{e^+} and k_{e^-}, publishing both $\mathrm{E}_{k_*^{\text{pub}}}(k_{\text{e}^+})$ and $\mathrm{E}_{k_*^{\text{pub}}}(k_{\text{e}^-})$, and then for each user publish $\mathrm{E}_{k_{\text{e}^+}} \left(\mathrm{E}_{k_i^{\text{pub}}}(x_i') \right)$ and $\mathrm{E}_{k_{\text{e}^-}}(y_i')$.

Thus, users can vote to end the protocol either by restoring access to all or by exposing each users' account destruction key. In practice, we might assume y_i' is in fact just x_i', the user's password. In this case, upon a successful decryption of k_{e^-}, there would be a race to take over the newly exposed accounts. The protest may rely on the action of griefers to quickly destroy or hijack the participants' accounts.

2.4.2 Individual escrow

It may be desirable for the crowd to be able to vote individually to unlock or destroy individual user accounts. This can be accomplished at some expense by using individual escrow keys $k^i_{\mathrm{e}+}$ and $k^i_{\mathrm{e}-}$ for each user and then publishing $\mathrm{E}_{k^i_{\mathrm{e}+}}\left(\mathrm{E}_{k_i^{\mathrm{Pub}}}(x'_i)\right)$ and $\mathrm{E}_{k^i_{\mathrm{e}-}}(y'_i)$ for each user in addition to the previous information.

This would enable a protest in which random accounts are destroyed regularly, giving the protest urgency and a limited time frame. The random account to destroy can be chosen by a distributed shared-randomness protocol, or P can specifically enable random deletion by randomising the destruction-key escrow, publishing $\mathrm{E}_{k^j_{\mathrm{e}-}}(i, y'_i)$ such that the protesters don't know which account i they are exposing if they reveal $k^j_{\mathrm{e}-}$. This can be considered a digital version of a crowd of street protesters exposed to random gunfire.

3 Limitations

3.1 Platform neutrality

This protest will not work if $\mathcal{O}$ chooses to interfere. In particular, $\mathcal{O}$ can prevent password changes, choose to remember old passwords and give users the option to back out of the protest using their old passwords, or can restore accounts that users have attempted to delete. The significance of this is that such a protest cannot be mounted against the service provider itself, or any party which the service provider has a vested interest in defending. In the real world, users have often used Facebook itself to protest changes to the site, which would not be possible in this setting.

3.2 Platform minority

Such a protest requires that the n participating members represent only a minority of the total population of the online service for two reasons. First, the service provider would probably interfere with the protest as described above if threatened with losing a majority of its members. Second, the effect of a large number of users losing their accounts might actually be less damaging, as they could collectively migrate to a different service. If only a small number participate in the protest, then the suicide option would be more damaging as participants would risk losing access to a still-popular service.

3.3 Value of identities

Most offline protests hold the central principle that all human lives are irreplaceable and of equal value. In an online protest, some identities may be much more valuable than others. In particular, some protesters may have new or little-used accounts which they are much more willing to lose than users with well-established accounts. The only realistic defense against this is to establish participation rules enforced by P during initialisation, such as a minimum age of accounts, and bar others from participating in the protest.

3.4 Infiltration

A protest could be hijacked by individuals not interested in the social aims of the movement, or indeed opponents of it. The hijacking could also take the form of a sybil attack, where one user with many accounts gains a large amount of voting power. Again, there appears to be little defence against this except to rely on P to vet candidates carefully. Cryptographically, it is desirable to use a robust threshold scheme, wherein a malicious user cannot poison the results of decryption by intentionally submitting an incorrect share. This ensures that malicious users cannot block legitimate votes to end a protest.

3.5 Central trust

The protest organiser P can interfere with the protest in various ways, and also hijack or destroy participants' accounts itself. It is difficult to de-couple P's dual roles of resetting users passwords and generating key shares, because we wish to ensure that only participants who have provided a valid account gain voting rights. Schemes do exist for threshold cryptography without a trusted dealer, but they would require giving all participants shares before a public key was available to escrow keys with. Such a scheme would yield great exposure to sybil attacks and infiltration because shares could be acquired prior to submitting valid credentials.

We appear to have little choice but to assume a benevolent initiator exists whom protestors are willing to trust in the role of P. Note that we have attempted to design the protocol so that P can go offline once the protocol has begun. This would allow P to in fact be a trusted elder or non-profit organisation who may not be particularly attached to a specific protest, but is trusted in general for organising social movements.

3.6 Splinter coalition

If serious disputes arise within the protest community, it might result in m users intentionally destroying the accounts of their opponents within the movement. This is particularly problematic in the "individual escrow" scheme whereby individuals can be targeted for destruction by design. Even in the basic scheme, a malicious coalition could secretly decrypt the escrow key, restore their own accounts, and refuse to share it with a targeted minority. Such a conspiracy would be inherently unstable, as it would require electing a new trusted leader and any defector could blow the whistle on its existence. If this is a significant concern, it may be necessary for P to maintain the master private key k_*^{priv} to respond to any allegations of conspiracy.

4 Conclusions

We have presented a novel protocol to enable digital protesters to engage with larger and graver issues. Ultimately, our proposal still relies on public perception and it is difficult to tell if an online protest can ever generate the same level of human compassion as physical suffering. Certainly, it would be disrespectful to compare the mortal sacrifices of history's great social advocates with the loss of a social networking profile.

However, there is evidence that online identities are becoming increasingly valuable, and indeed many users would be reluctant to give up access over a cause. Thus, we speculate that there may be some value to this approach in enabling a protest to demonstrate strong commitment by its members while remaining completely non-violent.

References

[1] Bridget M. Blodgett. And the Ringleaders Were Banned: An Examination of Protest in Virtual Worlds . In *C & T '09: Proceedings of the Fourth International Conference on Communities and Technologies*, pages 135–144, New York, NY, USA, 2009. ACM.

[2] Sasha Costanza-Chock. Mapping the Repertoire of Electronic Contention. In Andrew Opel and Donnalyn Pompper, editors, *Representing Resistance: Media, Civil Disobedience and the Global Justice Movement* , 2003.

[3] Dorothy Denning. Activism, Hacktivism, and Cyberterrorism: The Internet as a Tool for Influencing Foreign Policy. In *Networks and Netwars: The Future of Terror, Crime, and Militancy*, 2001.

[4] Dorothy Denning. A View of Cyberterrorism Five Years Later. In K. Himma, editor, *Internet Security: Hacking, Counterhacking, and Society*, 2007.

[5] Yvo Desmedt. Some Recent Research Aspects of Threshold Cryptography. In *ISW '97: Proceedings of the First International Workshop on Information Security*, pages 158–173, London, UK, 1998. Springer-Verlag.

[6] "Critical Art Ensemble". Electronic Civil Disobedience and Other Unpopular Ideas. Autonomedia, 1996.

[7] R. Kelly Garrett. Protest in an Information Society: A Review of Literature on Social Movements and New ICTs . In *Information, Communication and Society*, volume 9, pages 202–224, 2006.

[8] B. Kracher and K. Martin. A Moral Evaluation of Online Business Protest Tactics and Implications for Stakeholder Management . *Business and Society Review*, 2009.

[9] Seth Kreimer. Technologies of Protest: Insurgent Social Movements and the First Amendment in the Era of the Internet. *University of Pennsylvania Law Review*, 2001.

[10] Jill Lane. Digital Zapatistas. *The Drama Review*, 47, 2003.

[11] Ben Laurie and Richard Clayton. "Proof of Work" Proves Not to Work. *The Third Annual Workshop on the Economics of Inforomation Security*, 2004.

[12] Mark Manion and Abby Goodrum. Terrorism or Civil Disobedience: Toward a Hacktivist Ethic. *SIGCAS Comput. Soc.*, 30(2):14–19, 2000.

[13] Fiona McPhillips. Internet Activism: Towards a Framework for Emergent Democracy. *International Association for Development of the Information Society Journal on WWW/Internet*, 2006.

[14] Christina Neumayer and Celina Raffl. Facebook for Global Protest: The Potential and Limits of Social Software for Grassroots Activism . *Community Informatics Conference: ICTs for Social Inclusion: What is the Reality?*, 2008.

[15] Brett Rolfe. Building an Electronic Repertoire of Contention. In *Social Movement Studies*, volume 4, pages 65–74, 2005.

[16] Cass Sunstein. Echo Chambers. *Princeton University Press*, 2001.

[17] Julie Thomas. Ethics of Hacktivism. *Information Security Reading Room*, 2001.