CHAPTER 8

Ethical Issues Involving Computer Security: Hacking, Hacktivism, and Counterhacking

KENNETH EINAR HIMMA

This chapter considers whether and to what extent various types of unauthorized computer intrusions by private persons and groups (as opposed to state agents and agencies) are morally permissible;[1] this chapter does not cover other security-related issues, such as issues at the intersection of computer security and privacy, anonymity, and encryption.[2] The first section articulates a *prima facie* general case against these intrusions. The second considers intrusions motivated by malicious intentions and by certain benign intentions, such as the intent to expose security vulnerabilities. The third considers hacktivism, while the fourth considers counterhacking (or hackbacks).

Certain assumptions about "hacker" and related terms should be made explicit. Although these terms were once used to refer to accomplished programmers and their achievements, they are now used to refer to unauthorized computer intrusions and the persons who commit them. Thus construed, "hacking" is used, without moral judgment, to refer to acts in which one person gains unauthorized entry to the

---

Portions of this chapter appeared in my articles "Hacking as politically motivated digital civil disobedience: Is hacktivism morally justified?" and "The ethics of hacking back: active response to computer intrusions," which were originally published in Himma, K.E. (2007) *Internet Security: Hacking, Counterhacking, and Society.* Jones & Bartlett. I'm grateful to the publishers for permission to reprint those portions here.

[1]Public acts (i.e., those performed by state agencies) raise radically different issues. States are frequently permitted to do things that private individuals are not—such as incarcerating persons.

[2]For a discussion of these other issues, see Tavani, H. (2007). "The conceptual and moral landscape of computer security" In: Himma, K.E. (Ed.), *Internet Security: Hacking, Counterhacking and Society.* Jones & Bartlett, Sudbury, MA. In that essay, Tavani also differentiates ethical issues affecting three distinct aspects of computer security: data security, system security, and network security.

The Handbook of Information and Computer Ethics, Edited by Kenneth Einar Himma and Herman T. Tavani

computers of another person, and "hacker" is used to refer to someone who has committed such acts. Although some programmers bemoan the change in meaning, this chapter acquiesces to current usage.

8.1 THE *PRIMA FACIE* CASE AGAINST HACKING

At first glance, it might seem obvious that hacking is wrong. Although the more malicious of these acts involve serious wrongs because of the harm they cause, all are wrong because they constitute a digital trespass onto the property of another person. Unauthorized entry into some other person's computer seems not relevantly different than uninvited entry onto the land of another person. Real trespass is morally wrong, regardless of whether it results in harm, because it violates the owner's property right to control the uses to which her land is put and hence to exclude other people from its use. Similarly, digital trespass is wrong, regardless of whether it results in harm, because it violates the owner's property right to exclude other people from the use of her computer, which, like land, is physical (as opposed to intangible) property.

There are two problems with this argument. First, assuming that hacking is a species of trespass, it doesn't follow that all hacking is wrong because not all trespasses are wrong. It is permissible to trespass onto your land if doing so is the only way to capture a murderer fleeing the crime scene; committing a minor trespass is morally justified as the only way to secure the great good of stopping a killer. If hacking is trespass, then hacking necessary to secure some good that significantly outweighs the evil involved in trespass would also be justified.

Second, and more importantly, it is not clear that the concept of trespass properly applies to digital intrusions. The term "trespass" has largely been reserved—at least in moral usage—to refer to acts in which one person enters upon physical space owned by another, but a hacker is not in any *literal* sense *entering* upon a physical space owned by another person. Perhaps digital intrusion is more like using heat sensors to see what is going on inside a house, which is not usually characterized as "trespass," than like coming into the house without permission.

Even so, it seems clear that digital intrusions impinge upon legitimate interests of computer users. It seems clear, for example, that an unauthorized computer intrusion impinges upon the victim's property rights. Someone who gains access to my computer without my permission is appropriating a physical object in which I have a legitimate property interest; it is, after all, *my* computer—and I have, at the very least, a presumptive moral right to exclude other people from appropriating my computer.

If this is correct, then an unauthorized computer intrusion also impinges upon privacy rights. Someone who hacks into my computer without my permission gets access to something in which I have a legitimate expectation of privacy. If I may legitimately exclude others from my computer, then it is reasonable to regard my computer as a private space in which I can store sensitive information. Indeed, insofar as a computer user has a legitimate expectation of privacy in the contents stored on her computer, an unauthorized intrusion impinges upon the victim's privacy rights—regardless of whether there is, in fact, any sensitive information stored on that machine.

Moral rights are not, however, absolute. If I may trespass to capture a fleeing killer, then a person's right to property can be outweighed by more important rights when they conflict; property rights are weaker, for example, than the right to life. Similarly, a person's privacy rights can be outweighed by other more important interests; a person might, for example, be obligated to disclose sensitive information if needed to ensure another person's safety. Thus, the mere fact that a person has property and privacy rights in her computer does not imply that all unauthorized intrusions are impermissible.

Even so, the burden rests on the hacker to show that a particular intrusion is morally permissible. This will involve showing that any legitimate property or privacy interests are outweighed, as an ethical matter, by interests that can be secured only by committing the intrusion. Insofar as an intrusion involves causing damage to the files of the user, the intrusion can be justified only to the extent that it serves correspondingly greater interests.

Nevertheless, intrusions intended to cause harm out of *malice* are generally wrong. Although it is sometimes permissible to inflict harm on another person when necessary to secure a greater good, a malicious *intention* does not seek a greater good. This is not to say that it is necessarily wrong for one party to hack into another party's computer for the purpose of causing harm. Presumably, it is permissible to hack into the computers of known terrorists to delete files associated with a terrorist plot; given that innocent lives are at stake, this sort of digital harm seems clearly permissible. But these motivations are not *malicious*; the motivation here is to secure the greater good of saving lives, which justifies inflicting the comparatively minor harm of deleting files intended to advance an egregiously wrongful plot.

8.2 OVERCOMING THE *PRIMA FACIE* CASE: HACKING MOTIVATED BY BENIGN PURPOSES

Many hackers believe benign intrusions not calculated to cause damage can be justified on the strength of a variety of considerations. Such considerations include the social benefits resulting from such intrusions; speech rights requiring the free flow of content; and principles condemning waste. These arguments are considered in this section.[3]

8.2.1 The Social Benefits of Benign Intrusions

Hackers point out that benign intrusions have a number of social benefits. First, by gaining insight into the operations of existing networks, hackers develop knowledge that can be used to improve those networks. Second, the break-ins themselves call attention to security flaws that can be exploited by malicious hackers or, worse, terrorists. These are benefits that conduce to the public good and are thereby justified.

[3]I should acknowledge, at the very outset, that the discussion in this section has been deeply influenced by Spafford, E. (1992).

None of these benefits justifies benign intrusions. Even if we assume that privacy and property rights might sometimes yield to such utilitarian considerations, these social benefits can be achieved without infringing upon any moral rights. For example, hackers can develop these techniques and technologies in settings where the consent of all parties has been obtained. In cases where hackers seek entry to the machines of ordinary, noncommercial users, they can solicit the consent of other like-minded individuals to allow them to attempt to circumvent the relevant security measures. In cases where hackers seek entry to the machines of larger commercial users, they can seek employment at those firms or advise persons already employed at those firms. It is wrong to infringe privacy and property interests to achieve social benefits that can be achieved without infringing those rights.

There is, however, a deeper problem with this strategy of argument. If, as is commonly believed, the privacy and property interests of computer owners in their machines rise to the level of moral *rights*, then an appeal to social benefits cannot justify hacking. It is part of the very concept of a right that the infringement (as opposed to violation) of a right cannot be justified solely by an appeal to the desirable consequences of doing so. The mere fact that someone could do a lot of social good by stealing, say, a billion dollars from Bill Gates cannot justify stealing that sum if Gates has a property right to all of that money. As Ronald Dworkin famously puts the point, rights trump consequences.[4]

The social benefits argument, then, fails because it is the wrong kind of argument. The property and privacy rights computer owners have in their machines can justifiably be infringed by an unauthorized intrusion only if required to secure some more important right that outweighs those privacy and property rights. If rights trump consequences, then hackers must identify some stronger reason that justifies an intrusion: the appeal to social benefits, by itself, is insufficient to justify the intrusions.

8.2.2 Benign Intrusions as Preventing Waste

Hackers have also defended benign intrusions on the ground that they make use of computing resources that would otherwise go to waste. On this line of reasoning, it is morally permissible to do what is needed to prevent valuable resources from going to waste; benign hacking activity is justified on the strength of a moral principle that condemns squandering valuable resources in a world of scarcity in which there are far more human wants than resources to satisfy them.

This argument, unlike the social benefits argument, is the right kind of argument because it attempts to identify a moral principle that might limit other rights, like the right to property. Here it is crucial to note that rights are often limited by other moral principles; the right to life, for example, is limited by a moral principle that allows persons to kill if necessary to save their own lives from a culpable threat.

---

[4]See Dworkin, R. (1978). *Taking Rights Seriously.* Harvard University Press, Cambridge, MA.

Nevertheless, the argument fails. If one person has a property right in some object X, it is wrong for other persons to appropriate X without permission to prevent X from being wasted. As Spafford aptly puts this point:

> I am unable to think of any other item that someone may buy and maintain, only to have others claim a right to use it when it is idle. For instance, the thought of someone walking up to my expensive car and driving off in it simply because it is not currently being used is ludicrous. Likewise, because I am away at work, it is not proper to hold a party at my house because it is otherwise not being used. The related positions that unused computing capacity is a shared resource, and that my privately developed software belongs to everyone, are equally silly (and unethical) positions.

If it is wrong to appropriate someone's car without her permission to prevent waste, then there is no general moral principle that justifies infringing property rights to prevent waste and hence none that would justify hacking to prevent waste.

8.2.3 Benign Intrusions as Exercising the Right to a Free Flow of Content

This argument is grounded in the idea that the moral right to free expression entails that there should be no restrictions on the free flow of content; as this latter idea is sometimes put, information (or content generally) wants to be—or, better, ought to be—free. But if restrictions on the free flow of content are wrong in virtue of violating the right to free expression, then security measures designed to keep hackers out of networks violate their rights to free expression because they inhibit the free flow of content.

This argument also attempts to identify a moral principle that might limit other rights—indeed, one grounded in a putatively stronger right than privacy and property rights, namely the right of free expression. So strong is the right of free expression, on this analysis, that it entails a moral principle that would prevent any other right from permitting restrictions on the free flow of content.

It is, however, no more successful than the others in justifying hacking. The claim that there are no morally legitimate restrictions on the free flow of content precludes there being *any* right to informational privacy that entitles persons to exclude others from information in which they have a reasonable expectation of privacy; efforts that exclude others from information, by definition, impede the free flow of content. If we have any right to informational privacy (in, say, our medical records), as seems plausible, then the right to free expression permits restrictions on the free flow of content.

Further, the claim that there are no morally legitimate restrictions on the free flow of content is inconsistent with there being any moral intellectual property (IP) rights. Of course, moral IP rights might be much weaker than the right defined by IP law. But the idea that there are any moral IP rights is inconsistent with the claim that there are no morally legitimate restrictions on the free flow of content. If we have any moral IP right to exclude people from the contents of at least some of our creations, then the right of free expression permits restrictions on the flow of content.

But even if it were true that there are no legitimate restrictions on the free flow of content, it doesn't follow that people have *carte blanche* to get content any way possible. For example, it is clearly wrong to break into someone's house in order to gain information about what websites she visits. Even assuming that the right of free expression entails that other people are entitled to that information, there are limits on what persons can do *to exercise* that right. In ordinary circumstances, one cannot violate another person's property to exercise the right to free expression. If, as seems reasonable, hacking violates the legitimate property interests of a person in her computer (which is physical tangible property), then it is wrong regardless of whether the hacker is otherwise entitled to information on the victim's computer.

8.3 HACKTIVISM: HACKING AS POLITICALLY MOTIVATED ACTIVISM AND CIVIL DISOBEDIENCE

Recently a more plausible justification of hacking as protected free expression has emerged. According to this argument, attacks on government and corporate sites can be justified as a form of civil disobedience (CD) (Manion and Goodrum, 2000). Since CD is morally justifiable as a protest against injustice, it is permissible to commit digital intrusions to protest injustice. Insofar as it is permissible to stage a sit-in in a commercial or governmental building to protest, say, laws that violate human rights, it is permissible to intrude upon commercial or government networks to protest such laws. Thus, digital intrusions that would otherwise be morally objectionable are morally permissible if they are politically motivated acts of electronic CD—or "hacktivism," as such intrusions have come to be called.

8.3.1 CD and Morality

As a conceptual matter, CD involves (1) the open, (2) knowing (3) commission of some nonviolent act (4) that violates a law *L* (5) for the expressive purpose of protesting or calling attention to the injustice of *L*, some other law, or the legal system as a whole. An act need not target a law or system that is unjust as an objective moral matter to be properly characterized as "civil disobedience." It is enough that the actor is motivated by a belief that the law or system is unjust and that the act is contrived to protest it and call attention to its injustice. Since acts of CD are deliberately open so as to call attention to the putative injustice of the law or legal system, they are fairly characterized as "political expression."

It is tempting to think that acts of CD, as political expression, are morally justified as an exercise of the moral right to free expression. On this line of analysis, the right to free expression entails a right to express one's political views about the legitimacy of the law. Since the very point of CD is to call attention to the illegitimacy of the law, it is a morally justified exercise of the right to free expression.

This line of reasoning is problematic. First, the claim that X has a right to express *p* does not imply that expressing *p* is morally permissible. One might have a right to express all sorts of ideas it is morally wrong to express. For example, one might have a

right to express racist ideas even though expressing these ideas is wrong. Rights like the right to free expression are negative rights that are constituted by obligations on the part of other persons or entities. My right to free expression, for example, is constituted by an obligation on the part of others not to coercively interfere with my speech. But the claim that X has an obligation not to interfere with my saying *p* does not imply that my saying *p* is morally permissible; it just means that X should not use coercive means to prevent me from saying *p*.

Second, CD might be expressive, but it is primarily conduct. CD, by its nature, involves disobeying something with the status of law. It is one thing to *assert* a law is unjust; it is another thing to deliberately and openly behave in a manner that violates the law; the former is a pure speech act, while the latter is conduct. CD might be expressive conduct, but it is primarily *conduct* and secondarily expression.

Expressive conduct is subject to more stringent moral limits than those to which pure speech is subject. The reason for this has to do with the effects of these different kinds of act. As a general matter, pure speech acts are primarily calculated to affect only mental states. Speech acts intended to advance some view are calculated only to alter or reinforce the belief structure in the audience. In contrast, while conduct might frequently be *intended* to have only such effects, conduct tends to have effects on other important interests. Someone who expresses anger with you by hitting you not only affects your beliefs, but also causes you physical and emotional injury. Injury is just not a reasonably likely outcome from pure speech acts of just about any kind.

This is not to deny that violating the law might sometimes be morally permissible in certain circumstances. Legal and political philosophers are nearly unanimous in believing not only that there is no general moral obligation to obey the law, but also that there is no general moral obligation to obey the law of even reasonably just states; even reasonably just states, like morally wicked states, might sometimes enact legal content so wicked it does not generate a moral obligation to obey.

The circumstances in which one is permitted or obligated to disobey the law, however, will be comparatively rare in a morally legitimate democratic system with a body of law that is largely, though not perfectly, just for a number of reasons. First, citizens have alternative channels through which to express their political views in democratic systems with rights to free speech that can be exercised in a variety of ways—including blogs potentially reaching billions of people. Second, one of the virtues of democracy is that it affords each person an equal voice in determining what becomes law. Someone who violates democratically enacted law arrogates to herself a larger role than what she is entitled to in a democracy. It is *prima facie* problematic to circumvent legitimate democratic procedures in this way. Third, legitimate democratic states have latitude to enforce some unjust laws. Although some laws, like Jim Crow laws, are so unjust that it was wrong for states to coercively enforce them, a state may permissibly enforce some bad laws that do not reach some threshold level of injustice (e.g., tax laws that are not perfectly fair). But the state can be justified in coercively enforcing a law only insofar as a citizen is morally culpable in disobeying the law. This suggests that citizens are sometimes morally obligated to obey even bad laws.

Even so, the idea that there are limits on the scope of a legitimate state's permission to coercively enforce law indicates that CD is sometimes morally justified. In cases

where a legitimate state has enacted a sufficiently unjust law that falls outside the scope of its coercive authority, citizens have a qualified moral permission to disobey it. That is, in cases where the state is not justified in coercively enforcing a law, citizens may permissibly disobey that law because it does not give rise to any moral obligation to obey.

But this permission is qualified by a number of factors. The agent should be in cognitive possession of plausible justification for the position motivating the act of CD, and the position itself should be reasonable. Here it is important to note that the mental state of someone who commits an act of CD is not entirely unproblematic from a moral point of view. While such a person's *motivations* might be laudable, she will likely have another mental state that is not unproblematic from a moral point of view. Someone who commits an act of CD is usually acting on the strength of a conviction that is deeply contested in the society—and, indeed, one that is frequently a *minority* position.

This, by itself, can obviously be laudable in many circumstances. The courage to act on one's convictions and the willingness to sacrifice for them are both virtues. We encourage a child, for example, not to follow the crowd when it is wrong or foolish, knowing that such behavior will frequently result in unpleasant social consequences to the child, such as ridicule or ostracism. One who is willing to risk ridicule and ostracism in order to honor her moral convictions is courageous and deserves praise.

There are, however, moral limits on the costs one can impose on innocent third parties on the strength of an even a laudable motivation. After passage of a citizen initiative banning affirmative action by the state in Washington, protesters marched on a Washington highway in order to shut down traffic, something they succeeded in doing for hours. Those protesters *deliberately* caused significant inconvenience to other persons (many of whom voted against the initiative) after having their position rejected at the polls. Although their position might have been the correct one, their willingness to cause such inconvenience to others on the strength of a view that might, or might not, have been particularly well reasoned is morally problematic—even if, all things considered, their conduct was morally permissible.

The mental state of someone who deliberately imposes detriment on innocent third parties on the strength of a moral conviction that lacks adequate epistemic support is morally problematic in at least two possibly related ways. First, it evinces disregard for the interests of innocent third parties, a failure to appreciate the importance of other people. Second, it evinces an arrogant judgment about the importance and reliability of one's own judgments. It seems, at the very least, arrogant for one person to deliberately subject another to a risk of harm on the strength of an idea that lacks adequate support.[5]

[5]This primarily applies to individuals; the state is in a somewhat different position because, in many cases, it cannot avoid taking a position on a contested issue by refraining from acting. If it refrains from prohibiting abortion, for example, the absence of a prohibition presupposes (at least to the extent that we presume the state is trying to do what is morally legitimate) that abortion does not result in murder. If it prohibits abortion, the prohibition presupposes either that a woman does not have a privacy right in her body or that abortion results in murder. Citizens are rarely in such a position.

One sign that a moral conviction lacks adequate epistemic support is that it is deeply contested among open-minded, reasonable persons of conscience in the culture. The idea that there are many open-minded reasonable persons of conscience on both sides of an issue suggests that both positions are reasonable in the sense that they are backed by good reasons that lack an adequate rebuttal. Insofar as a disagreement is *reasonable* in this sense, neither side can claim to have fully adequate support.[6]

Agents must also be willing to accept responsibility under the law for their acts of CD. Willingness to accept responsibility goes beyond merely openly defying the law; one might openly defy a law but attempt to evade apprehension by the police by, say, leaving the country. Intuitively, there is a world of difference, for example, between someone who defaces a billboard in front of 15 police officers to protest its content and someone who does so in a clandestine manner hoping to avoid detection. One seems fairly characterized as vandalism while the other, even if ultimately unjustified, does not.

This last factor is especially important in evaluating hacktivism. Many theorists worry that it can be difficult to distinguish hacktivism from cyberterrorism. Huschle (2002) argues that hacktivists should make it a point to accept responsibility for their actions precisely to ensure that their acts are not mistaken for cyberterrorism, which could cause much more disruption than was intended to result from their acts. As Manion and Goodrum (2000) put the point:

> The justification of hacktivism entails demonstrating that its practitioners are neither "crackers"—those who break into systems for profit or vandalism—nor are they cyberterrorists—those who use computer technology with the intention of causing grave harm such as loss of life, severe economic losses, or destruction of critical infrastructure. Hacktivism must be shown to be ethically motivated (pp. 15–16).

The acceptance of responsibility and the legal consequences of disobedience signals that the act is motivated by a principled stand, a feature that operates to legitimize these acts. Moreover, the willingness of the agent to accept responsibility signals that the breach of the public peace is exceptional rather than part of a general pattern of misconduct and hence need not give rise to the feelings of vulnerability and insecurity to which breaches of the public peace typically give rise.

The foregoing discussion suggests a useful framework for evaluating acts of CD that weigh the moral benefits and costs. The following considerations weigh in favor of finding that an act of CD in a legitimate democratic state is morally permissible. First, the act is committed openly by properly motivated persons willing to accept responsibility for the act. Second, the position is a plausible one in play among open-minded, reasonable persons in the relevant community. Third, persons committing an act of CD are in possession of a thoughtful justification for both the position and

[6]Not all disagreement is reasonable, of course. It seems clear, for example, that persons who disagreed in the 1960s and 1970s with the idea that race-based segregation is wrong lacked even minimal support for a position that had been all but conclusively refuted by that juncture. Sometimes there are a lot of unreasonable, narrow-minded persons who simply refuse to see the light.

the act. Fourth, the act does not result in significant damage to the interests of innocent third parties. Fifth, the act is reasonably calculated to stimulate and advance debate on the issue.

In contrast, the following considerations weigh in favor of finding that an act of CD against an otherwise legitimate state is morally wrong. First, the act is not properly motivated or committed openly by persons willing to accept responsibility. Second, the position is implausible and not in play among most thoughtful open-minded persons in the community. Third, the people who have committed an act of CD lack a thoughtful justification for the position or the act. Fourth, the act results in significant harm to innocent third parties. Fifth, the act is not reasonably calculated to stimulate or advance debate on the issue.

The civil rights sit-ins of the 1960s are paradigms for justified acts of CD under the above framework. Someone who refuses to leave segregated lunch counters until police arrive to remove her is clearly committing an open act and is willing to accept the consequences. The view that segregation is wrong was not only in play among open-minded, reasonable persons of conscience, but had pretty much won the day by the time the mid-1960s arrived. The people who committed these acts justified them by reference to a principle of equality that open-minded, reasonable persons of conscience in the culture had nearly universally accepted. Lunch counter sit-ins had significant effects only on the owners who wrongly implemented policies of segregating blacks and whites. These sit-ins helped to call attention to the ongoing racial injustices in the southern United States.

In contrast, acts of vandalism by anarchists during the 1999 World Trade Organization (WTO) protests in Seattle were not justified acts of CD under this framework. Anarchists who broke windows typically fled the scene as soon as the police arrived. Anarchism is not in play among reasonable, open-minded persons in the community. If televised interviews with many of them were any indication, they generally lacked a thoughtful justification for their views; most of them I saw were strikingly inarticulate. The cost of replacing a large plate-glass storefront window is in excess of $10,000—a morally significant cost to innocent store owners. These acts of vandalism tend to alienate people and entrench them further in their opposition to anarchism, rather than provoke reasoned discussion.

8.3.2 What is Hacktivism?

For our purposes, "hacktivism" can be defined as "the commission of an unauthorized digital intrusion for the purpose of expressing a political or moral position." *Qua* digital act, hacktivism is nonviolent in nature. *Qua* activism, hacktivism does not seek to achieve its political purposes, unlike terrorism, by inspiring terror among the population; it attempts to achieve these purposes by stimulating discussion and debate. Hacktivism is thus conceptually distinct from cyberterrorism—though the boundaries, as we will see, sometimes seem to blur. Hacktivism is distinct from other forms of benign hacking (e.g., motivated by a desire for knowledge) in that it is motivated by the laudable desire to protest injustice.

Not all digital activism counts as hacktivism or CD. Posting a Web site in the United States with a petition to end the war in Iraq would be a form of digital activism, on this definition, but would not count as hacktivism because it does not involve an unauthorized digital intrusion. Nor, for that matter, would such an act count as an act of electronic CD because the posting of such content online breaks no laws; CD necessarily involves violating a valid law.

In contrast, the following count as both hacktivism and CD: (1) a denial-of-service (DoS) attack launched against the WTO Web site to protest WTO policies;[7] (2) the altering of the content of a government Web site to express outrage over some policy of that government;[8] and (3) the unauthorized redirection of traffic intended for a KKK Web site to Hatewatch.[9] Each of these acts would involve some unauthorized digital intrusion and hence would, since presumably intended as a piece of political activism, count as hacktivism.

8.3.3 Is Hacktivism Morally Justified as CD?

The issue of whether hacktivism is justified CD must be addressed on a case-by-case basis. Some hacktivists, for example, make no attempt to conceal their identity and accept responsibility, while others conceal their identities to evade detection. Some acts do not involve significant damage to innocent third parties (e.g., defacing a governmental Web site to protest its policies), while others do [e.g., shutting down commercial Web sites with distributed denial-of-service (DDoS) attacks]. Open acts of hacktivism that do not impact innocent third parties have a different moral quality than clandestine acts that harm innocent third parties.

One of the key issues in evaluating whether an act of hacktivism is morally justified is the extent to which the act harms the interests of innocent third parties. In thinking about this issue, it is important to reiterate that the context being assumed here is a morally legitimate democratic system that protects the right of free expression and thus affords persons a variety of avenues for expressing their views that do not impact the interests of innocent third parties.

How much harm is caused depends on whether the target is a public, private, commercial, or noncommercial entity. Attacks on public noncommercial, purely informative Web sites, for example, tend to cause less damage than attacks on private, commercial Web sites. The reason is that attacks on commercial Web sites can result in significant business losses passed on to consumers in the form of higher prices or to employees in the form of layoffs. If the information on a public Web site is nonessential

---

[7]In 1999, the Electrohippies attacked a WTO Web site for such reasons. For a summary of notable hacker attacks, see "Timeline of hacker history," *Wikipedia*; available at http://en.wikipedia.org/wiki/Infamous_Hacks.

[8]In 1996, hackers changed the content of the Department of Justice Web site, replacing "Justice" with "Injustice."

[9]Anonymous hackers did exactly this in 1999. Intriguingly, a Hatewatch press release characterized the act as "vandalism." See Hatewatch Press Release: Activism versus Hacktivism, September 4, 1999. Available at http://archives.openflows.org/hacktivism/hacktivism01048.html.

(i.e., unrelated to vital interests), an attack on that Web site is likely to result in nothing more serious than inconvenience to citizens who are not able to access that information.[10]

This should not be taken to suggest that hacktivist intrusions upon public entities *cannot* result in significant harm to third parties. One can conceive of a depressingly large variety of acts that might very well cause significant damage to innocent third parties. A digital attack on a public hospital server might very well result in deaths. Of course, these more serious acts are probably not motivated by expressive purposes and, if so, would not count either as CD or as hacktivism as these notions are defined here.

Acts of hacktivism directed at private individuals can also have morally significant effects. A DoS attack, for example, that effectively denies access to a citizen's Web site can impact her moral rights. A DoS attack on a citizen's site impacts her ability to express her views and hence infringes her moral right to free expression. An attempt to gain access to files on a citizen's computer impacts her rights to privacy, as well as her property rights in her computer.

How much harm is done also depends on the nature of the attack.[11] As a general matter, some digital attacks are less likely to cause harm than others. Defacement of a Web site—or "E-graffiti" as sympathetic theorists sometimes call it—seems far less likely to cause significant harm than attacks that simply deny access to a Web site. Changing "Department of Justice" on a government Web site to "Department of Injustice" is not likely to result in significant harm to third-party interests. At most, it will cause embarrassment to the government agency running the site.

This should not, however, be taken to suggest that defacement of a Web site can never result in significant damage to innocent third parties. Publishing sensitive information about individuals, like social security numbers, as "E-graffiti" on a government Web site could obviously result in significant damage to those individuals. As is true of physical graffiti, one must look to the specific circumstances to evaluate the damage caused by digital graffiti to ensure an accurate assessment.

Nevertheless, it is reasonable to think that, as a general matter, DoS and DDoS attacks are likely to cause more damage, other things being equal, than defacement of Web sites. These attacks are calculated to deny access of third parties to the content of a Web site, effectively shutting it down by overwhelming the server with sham requests for information. While there are undoubtedly exceptions to any generalizations about the comparative harm caused by defacement of Web sites and DoS attacks, it seems reasonable to think that shutting down a Web site is a more harmful act than merely defacing it. For this reason, DoS attack will be harder to justify, as a general matter, as permissible electronic CD than defacement.

---

[10]The Electrohippies justify attacks on various public Web sites precisely on such grounds: "Neither the Whitehouse nor 10 Downing Street Web site are [sic] essential services. For the most part they merely distribute the fallacious justifications in Iraq, as well as trying to promote the image of the two prime movers behind war in Iraq: Messrs. Bush and Blair" (Electrohippies, 2003). The idea here is that the harm caused by attacks that ultimately deny access to public Web sites that are not providing essential services results in no significant harm to innocent third parties.

[11]For a helpful discussion of various tactics, see Auty (2004) . My discussion in this and the last section owes an obvious debt to Auty's discussion.

Indeed, a coordinated and sustained DDoS attack on the largest commercial Web sites could result in an economic downturn that affects millions of people. Al Qaeda is exploring the possibility of large-scale cyberattacks on public and commercial networks precisely because a large enough attack might suffice to weaken confidence in E-commerce to such an extent as to precipitate a recession—or worse. Here it is worth noting that an increase of the unemployment rate in the United States from 5% to 6% means the loss of approximately one and a half million jobs—a consequence of great moral significance.

Another important factor in evaluating an act of CD is that the persons committing the act are willing to accept responsibility for those acts. Manion and Goodrum (2000), for example, assert that "willingness (of participants) to accept personal responsibility for outcome of actions" is a necessary, though not sufficient, condition for the justification of an act of CD: "In order for hacking to qualify as an act of civil disobedience, hackers must be clearly motivated by ethical concerns, be nonviolent, and be ready to accept the repercussions of their actions" (p. 15).

There is a difference between claiming responsibility for an act and being willing to accept the legal consequences of that act. One can claim responsibility without coming forward to accept the legal consequences of one's act. One can do this by giving some sort of pseudonym instead of one's real name or by attributing the act to a group that protects the names of its members. Although such a claim of responsibility signals an ethical motivation, this is not tantamount to being willing to accept responsibility.

The heroic civil rights activists of the 1960s who staged sit-ins went beyond merely claiming responsibility; they accepted, even invited, prosecution. It was part of their strategy to call attention to the injustice of Jim Crow laws in the South by voluntarily subjecting themselves to prosecution under those very laws. These courageous activists did not anonymously claim responsibility for the sit-ins from a safe distance: they would continue the protests until the police arrived to arrest them.

Some noteworthy hacktivists evince a similar willingness to accept responsibility for their actions. As Manion and Goodrum (2000) observe:

> Examined in the light, the hack by Eugene Kashpureff clearly constitutes an act of civil disobedience. Kashpureff usurped traffic from InterNIC to protest domain name policy. He did this nonanonymously and went to jail as a result (p. 15).

But this is the exception and not the rule. There are a variety of hacktivist groups, including Electrohippies, MilwOrm, and Electronic Disturbance Theatre, but these groups typically claim responsibility for acts *as a group* without disclosing the identities of any members. For example, MilwOrm and another group claimed responsibility for the defacement of approximately 300 Web sites (they replaced the existing content with a statement against nuclear weapons and a photograph of a mushroom cloud), but did not disclose the identities of members who belong to the group. Hacktivists typically attempt to conceal their identities to avoid exposure to prosecution—even when claiming responsibility.

Anonymous hacktivist attacks impose significant costs on social well-being. First, such attacks, regardless of motivation, contribute to an increasing sense of anxiety

among the population about the security of the Internet, which has become increasingly vital to economic and other important interests. Second, these attacks require an expenditure of valuable resources, which could be allocated in more productive ways, to protecting computers against intrusions—costs that are passed on to consumers.

In any event, it is worth noting that terrorists typically claim responsibility as a group, but attempt to evade the consequences of their actions by concealing their identities and locations. It is important, of course, not to make too much of this similarity: terrorists deliberately attempt to cause grievous harm to innocent people while hacktivists do not. The point, however, is merely to illustrate that there is a morally significant difference between claiming responsibility and accepting responsibility. Accepting responsibility is, other things being equal, needed to justify an act of hacktivism.

A third factor to consider is that the motivating agenda behind electronic CD, other things being equal, is not as transparent as the motivating agenda behind ordinary CD. Whereas the protesters who shut down the Washington state highway carried signs and alerted the press they were protesting a specific measure, the point of many putative acts of hacktivism is not clear. A DDoS attack, for example, directed against Amazon.com could mean any number of things—some of which have nothing to do with expressing a political view (e.g., a recently discharged employee might be taking revenge for her dismissal). The absence of a clear message is problematic from a moral standpoint.

Acts of hacktivism are frequently motivated to protest the violation of human rights by oppressive nondemocratic regimes and are directed at servers maintained and owned by governmental entities in those regimes. It is worth noting that, strictly speaking, many such acts will not count as CD. The reason is that many of these attacks will be from people who live outside the repressive regime and are not subject to the legal consequences within the regime. But insofar as these legal consequences are draconian and drastically out of proportion to what is morally appropriate, acceptance of responsibility is not necessary for such acts to be justified. Accordingly, these attacks that originate from outside the target nation might be justified hacktivism, but they will not be justified *as CD*.

Other features suggesting that such acts are justified as follows. First, the primary impact of such acts is on the parties culpable for committing violations of human rights. Defacing a governmental Web site that does not provide essential services or information is not likely to have any significant effects on innocent citizens. Second, the targeted regimes do not respect a right of free expression and forcefully repress political dissent. It is reasonable to think that the moral calculus of CD is considerably different in states that systematically deny citizens the opportunity to express dissent without fear of reprisal. Third, such acts of CD are frequently successful in calling attention to the injustice and stimulating debate. Finally, the position is probably a majority position among people in this culture and worldwide. It is fairly clear that, in Western cultures, support for universal human rights is, far and away, a majority position. But it is also reasonable to think that such support is also a majority position in non-Western cultures. In nations where citizens are

denied human rights, those citizens frequently demand them. When liberated from oppressive regimes, moreover, citizens tend to behave in ways that were suppressed under those regimes. Women, for example, in Afghanistan adopted a Western style of dress and rejected the oppressive burqa after the Taliban was removed from power. People almost universally want speech rights, equality, and a right to be free from torture or political persecution.

But, unlike the human rights agenda, other positions commonly motivating hacktivism are fairly characterized as fringe positions not generally in play among thoughtful, open-minded members of the community. Consider, for example, the main tenets of the "hacker ethic" as summarized by Levy (1984):

(1) Access to computers should be unlimited and total.
(2) All information should be free.
(3) Mistrust authority—promote decentralization.
(4) Hackers should be judged by their hacking, not by bogus criteria such as degrees, age, race, or position.
(5) You create art and beauty on a computer.
(6) Computers can change your life for the better.

Although tenets 4 through 6 are largely uncontroversial (and so obvious that they do not need to be stated), these are not the tenets that motivate acts of hacktivism. The tenets that are most likely to motivate acts of hacktivism are the first three tenets.

It is hard to know what to say about tenet 3 beyond pointing out that it is overly general (Should all doctors be mistrusted? Always?); however, tenets 1 and 2 are clearly fringe positions not in play among open-minded, thoughtful people. Tenet 1 implies that people have no property rights in their own computers and hence may not permissibly exclude others from their machines—an implausible position that, consistently applied to other forms of property, would vitiate ownership in homes and automobiles. Tenet 2 implies that people have no privacy rights in highly intimate information about themselves. Although many people are rightly rethinking their positions about information ethics in response to the new technologies, tenets 1 and 2 are too strong to be plausible because they are inconsistent with bedrock views about privacy and property rights. For this reason, neither is in play among open-minded, reasonable persons in the community. This operates against thinking hacktivism expressing the hacker ethic is justified.

Nevertheless, it is not enough, according to the framework described above, that an act of hacktivism is motivated by a plausible position in play among thoughtful, reasonably conscientious persons; it is a necessary condition for an act of hacktivism to be justified that the actor be in cognitive possession of a reasonably plausible justification for that position.

As a general matter, hacktivists give little reason to think they are in possession of a reasoned justification supporting the positions they take. Occasionally, they will articulate their position with some sort of slogan, but rarely provide the position with

the critical support it needs. Consider, for example, Manion's and Goodrum's (2000) discussion of one such motivation:

> In order to determine the motivations of hacktivists, one place to look is what hacktivists *themselves* say in their motivation . . .. In June, 1998 the hacktivist group "MilwOrm" hacked India's Bhabha Atomic Research Centre to protest against recent nuclear tests. Later, in July of that year, "MilwOrm" and the group "Astray Lumberjacks," orchestrated an unprecedented mass hack of more than 300 sites around the world, replacing web pages with an antinuclear statements(sic) and images of mushroom clouds. Not surprisingly, the published slogan of MilwOrm is "Putting the power back in the hands of people" (Manion and Goodrum, 2000, p. 16).

One should say much more by way of justification for hacking 300 sites than just a vague slogan like "Putting the power back in the hands of people." The victims of such an attack, as well as the public whose peace has been breached, have a right to know exactly what position is motivating the attack and why anyone should think it is a plausible position.

The foregoing argument should not, of course, be construed to condemn all acts of hacktivism. Nothing in the foregoing argument would condemn narrowly targeted acts of electronic CD properly motivated and justified by a well-articulated plausible position that do not result in significant harm to innocent third parties. Acts of hacktivism that have these properties might be justified by the right to free expression—though, again, it bears emphasizing here that such acts will be much harder to justify in societies with morally legitimate legal systems.

But, as a general matter, hacktivists have not done what they should to make sure their acts are unproblematic from a moral standpoint. In their zealousness to advance their moral causes, they have committed acts that seem more problematic from a moral point of view than the positions they seek to attack. If, as Manion and Goodrum (2000) suggest, hacktivists have been misunderstood by mainstream media and theorists, they have only themselves to blame.

8.4 HACKING BACK: ACTIVE RESPONSE TO COMPUTER INTRUSIONS

Victims of digital intrusions are increasingly responding with a variety of "active responses." Some are intended to inflict the same kind of harm on the attacker as the attack is intended to have on the victim. Conxion, for example, overloaded the network from which the Electrohippies staged a DoS attack by redirecting the incoming packets back to the network instead of dropping them at the router.[12] Some, however, are not intended to inflict harm on the attacker's network. The point of a traceback is to identify the parties responsible for the intrusion by tracing its path back to the source.

---

[12]See Radcliff, D. (2000). Should you strike back? *ComputerWorld.* Available from http://www.computerworld.com/governmenttopics/government/legalissues/story/0,10801,53869,00.html.

This section considers whether and to what extent it is morally permissible for private parties to adopt these active responses.

8.4.1 The Active Response Spectrum

The term "active response" is intended to pick out digital intrusions that come in response to a hacker's intrusion and are intended to counter it; these responses are sometimes called "counterhacking" or "hacking back." As such, active response measures have the following characteristics. First, they are digitally based; assaulting someone who is committing a digital trespass is not active response. Second, they are implemented after detection of an intrusion and are intended to counter it by achieving investigative, defensive, or punitive purposes. Third, they are noncooperative in that they are implemented without the consent of at least one of the parties involved in or affected by the intrusion. Finally, they have causal impacts on remote systems (i.e., those owned or controlled by some other person).

"Benign" responses involve causal interaction with remote systems outside the victim's network, but are neither intended nor reasonably likely to damage those systems. One example of a benign response is a traceback. As noted above, tracebacks attempt to identify the parties responsible for an digital intrusion by following its path in reverse; they causally impact remote systems but without damaging them.

"Aggressive" responses are those calculated to interfere with the availability, integrity, confidentiality, or authenticity of remote systems. Aggressive measures are those intended or highly likely to result in something that the target would regard as harm or damage. An example of an aggressive response is a DoS counterattack of the sort launched by Conxion against Electrohippies.

8.4.2 Relevant Moral Principles

8.4.2.1 A Principle Allowing Force in Defense of Self and Others It is generally accepted that a person has a moral right to use proportional force when necessary to defend against an attack. If, for example, A starts shooting at B without provocation and B cannot save her own life without shooting back at A, it is permissible for B to shoot at A. The first principle considered here, then, can be stated as follows:

> **The Defense Principle:** It is morally permissible for one person to use force to defend herself or other innocent persons against an attack provided that (1) such force is proportional to the force used in the attack or threat; (2) such force is necessary either to repel the attack or threat or to prevent it from resulting in harm; and (3) such force is directed only at persons who are the immediate source of the attack or threat.

Although the term "force" has traditionally been used to refer to violent physical attacks in which one person attempts to inflict physical harm on another person, it is construed here as applying to both physical *and* digital attacks.

Each element of the Defense Principle states a necessary condition for the justified use of force. First, the Defense Principle justifies the use of no more force than is proportional to the attack. Second, force must be necessary in the sense that the victim cannot either stop the attack or prevent further harm to herself without resorting to its use. Third, the Defense Principle will justify the use of force only against the direct sources of the attack. In limited cases, this might permit the use of force against innocent persons. Many people have the intuition that a person may direct force against an attacker known to be insane and hence not responsible for her actions. Nevertheless, the Defense Principle will never justify directing force against an innocent bystander. Under no circumstances, then, would it allow a person to defend against an attack by interposing an innocent bystander between herself and the attacker.

8.4.2.2 A Principle Allowing Otherwise Wrongful Acts to Secure Greater Moral Good

It is also generally accepted that morality allows the infringement (as opposed to violation) of an innocent person's rights when it is necessary to secure a significantly greater good.[13] For example, if A must enter onto the property of B without her permission to stop a murderer from escaping, it is morally permissible for A to do so. Though such an act constitutes a trespass and hence infringes B's property rights, it does not violate B's property rights because it is morally justified. This suggests a second general principle relevant in evaluating an active response:

> **The Necessity Principle:** It is morally permissible for one person A to infringe a right ρ of a person B if and only if (1) A's infringing of ρ would result in great moral value; (2) the good that is protected by ρ is significantly less valuable, morally speaking, than the good A can bring about by infringing ρ; (3) there is no other way for A to bring about this moral value that does not involve infringing ρ; and (4) A's attitude toward B's rights is otherwise properly respectful.

As construed here, the Necessity Principle applies in the context of physical and digital attacks and hence potentially justifies the use of physical or digital force that would ordinarily be impermissible.

Each element of the Necessity Principle states a necessary condition for being justified in doing something that would otherwise be wrong. First, the act is not justified unless it results in a significantly greater good than the interest infringed by the act. Second, the act is justified only if there is no other way to bring about the greater good. Third, the act must be performed with an otherwise respectful attitude.

The Necessity Principle augments the Defense Principle by allowing some action that would infringe the rights of even innocent bystanders: the Necessity Principle seems to allow one person A to infringe the right of an innocent bystander B if

[13]By definition, to say that a right has been "infringed" is to say only that someone has acted in a way that is inconsistent with the holder's interest in that right; strictly speaking, then, the claim that a right has been infringed is a purely descriptive claim that connotes no moral judgment as to whether or not the infringement is wrong. In contrast, to say that a right has been "violated" is to say that the right has been infringed by some act and that the relevant act is morally wrong. Accordingly, it is a conceptual truth that it can be permissible for an individual or entity to infringe a right, but it cannot be permissible to violate a right.

necessary to defend A or some other person from a culpable attack that would result in a significantly greater harm than results from infringing B's right. But insofar as the Necessity Principle requires the achievement of a *significantly* greater good, it will not allow a person to direct at an innocent bystander force that is proportional to the force of the attack.

8.4.2.3 Two Nonstarters: Retaliation and Punishment It might be thought that victims of an attack have a moral right to retaliate against or punish their attackers by inflicting a proportional harm on their attackers. If, for example, A hits B in the face and then turns and runs away in an obvious attempt to escape, it is morally permissible, on this view, for B to catch A and hit him back in the face; such a measure is permissible either as retaliation or as punishment. Applied to the present context, such a principle would permit the victim of a digital attack to counterattack as a means of "evening the score."

Active response cannot be justified as retaliation. The act of inflicting injury on another person for no other reason than to even the score is "revenge," and revenge is generally regarded as morally wrong because it is no part of the concept of revenge that harm be inflicted to give a person his just deserts. From the standpoint of someone who is retaliating, the point of the retaliatory act is not to restore the balance of justice after it has been disturbed by a wrongful act. Rather, the point is simply to even the score: he did this to me, so I did it back to him. As far as ordinary intuitions go, morality does not allow the infliction of harm on another person without regard for whether that harm is deserved or serves some greater purpose than satisfying a desire for vengeance.

Nor can active response be justified as punishment. In a society with a morally legitimate government, it is morally impermissible for *private citizens* to punish wrongdoing. Mainstream political theorists are nearly unanimous in holding that it is the province of a legitimate government – and not of private persons – to punish wrongdoers after they are found guilty in a fair trial with just procedures. Indeed, vigilantism is universally condemned as morally wrong. Both lines of argument are nonstarters.

8.4.3 An Evidentiary Restriction for Justifiably Acting Under Ethical Principles

As was noted in the discussion of hacktivism, a person must have adequate reason to believe she is justified in acting under a moral principle to be justified in acting under that principle. There is thus another general principle relevant with respect to evaluating an active response—one that is epistemic in character:

> **The Evidentiary Principle:** It is morally permissible for one person A to take action under a moral principle P only if A has adequate reason for thinking that all of P's application-conditions are satisfied.

The Evidentiary Principle implies that one has a duty to ensure that one is epistemically justified in acting under the relevant moral principle. If one person

A takes aggressive action against another person B without sufficient reason to believe the application-conditions of the relevant moral principles have been satisfied, A commits a wrong against B.

Accordingly, the victim of a digital attack can permissibly adopt active response only if she has adequate reason to think the application-conditions of one of the relevant principles are satisfied. Under the Defense Principle, she must have adequate reason to believe that (1) whatever force is employed is proportional to the force used in the attack; (2) such force is necessary either to repel the attack or to prevent it from resulting in harm of some kind; and (3) such force is directed only at persons immediately responsible for the attack. Under the Necessity Principle, she must have adequate reason to believe that (1) the relevant moral value significantly outweighs the relevant moral disvalue; (2) there is no other way to achieve the greater moral good than to do A; and (3) doing A will succeed in achieving the greater moral good.

8.4.4 Evaluating Active Response Under the Relevant Principles

It is important to realize that the risk that active responses will impact innocent persons and their machines is not purely "theoretical." Sophisticated attackers usually conceal their identities by staging attacks from innocent machines that have been compromised through a variety of mechanisms. Most active responses will have to be directed, in part, at the agent machines used to stage the attack. Accordingly, it is not just *possible* that any efficacious response will impact innocent persons, it is nearly inevitable—something that anyone sophisticated enough to adopt an active response is fairly presumed to realize.

Given that innocent persons enjoy a general (though not unlimited) moral immunity against forceful attack, the likelihood of impacting innocent persons with an active response is of special ethical concern. For this reason, the impacts of active responses on innocent parties will occupy a central role in evaluating those responses.

8.4.4.1 Aggressive Measures As a general matter, aggressive active defense cannot be justified by the Defense Principle. Consider, again, Conxion's response to the Electrohippies DoS attack. Instead of simply dropping the incoming packets at the router, Conxion sent those packets back to the Electrohippies' server, overwhelming it. Since dropping the packets at the router would have ended the harmful effects of the attack, Conxion's response was not "necessary" and hence not justified under the Defense Principle.

Additional issues are raised by aggressive response to attacks staged from innocent agent machines. Since the identity of the culpable attacker is unknown in such cases, any aggressive response will invariably be directed at the innocent agents compromised by the attacker, which compounds the harms done to the owners of those machines. Even if it is permissible to use force against innocent attackers, the owners of those machines are not really "attackers" in the sense that an insane person who assaults another person is. If those owners are really "bystanders," the Defense Principle will not allow aggressive response to attacks in which it is evident that the attacker's identity has been concealed.

Moreover, aggressive response will not be justified under the Defense Principle if it is not *necessary* to prevent the harm or to stop the attack. If there is any nonaggressive way for the victim to avoid the attack or the damage caused by it, then an aggressive response cannot be justified under the Defense Principle. This, however, does not mean that the victim is obligated to escape the attack by any means possible. Although it is always possible to escape a digital attack by taking the target offline, such measures can result in significant damage (if, e.g., the target is a web-based business) that the Defense Principle does not require victims to accept. Victims have a duty to escape attacks only insofar as this can be done without incurring injuries that are comparable to those caused by the attack itself.

Aggressive response is problematic under the Necessity Principle for a different reason. The Necessity Principle allows acts that would otherwise be wrong if they are necessary to achieve a significantly greater moral good. Even if we assume that an aggressive response is clearly necessary to achieve the moral good of preventing the damage caused by an attack and that this good significantly outweighs the harms done to the owners of the agent machines, there is an evidentiary problem: for all we can know, an aggressive response might result in unpredictable harms that outweigh the relevant moral goods.

The problem here arises because machines can be linked via a network to one another in a variety of unpredictable ways, making it impossible to identify all the harmful effects of an aggressive response in advance. Suppose, for example, that an attacker compromises machines on a university network linked to a university hospital. If hospital machines performing a life-saving function are linked to the network, an aggressive response against that network might result in a loss of human life. Even worse, suppose an attacker compromises machines used by one nation's government to attack private machines in another nation. If the two nations are hostile toward each other, an aggressive response by the private victim could raise international tensions—a particularly chilling prospect if the two nations are nuclear powers.

The point is not that we have reason to think that these scenarios are likely; rather, it is that we do not have any reliable way to determine how likely they are. A victim contemplating an aggressive response has no reliable way to estimate the probabilities of such scenarios in the short time available to him or her. Since the victim cannot reliably assess these probabilities, she lacks adequate reason to think that the application-conditions of the Necessity Principle are satisfied. Thus, under the Evidentiary Principle, she may not justifiably adopt aggressive measures under this principle.

8.4.4.2 Benign Measures Benign measures are typically concerned with identifying culpable attackers (e.g., tracebacks) and are neither intended nor obviously likely to result in physical damage to affected machines. Even so, benign responses are problematic insofar as they causally impact remote machines. Of course, this does not pose any obvious moral problems when the remote machines are located within the victim's network or when the victim has permission to impact these machines. But unauthorized effects on innocent agent

machines are presumptively problematic since they infringe the property rights of an innocent person.

Benign responses do not defend against attacks and hence cannot be justified by the Defense Principle, but they seem to secure an important moral good under the Necessity Principle. Criminal attacks are regarded as offenses against the general public because they violate the legitimate expectations of the public and thereby breach the peace. Like the victim, the public has a compelling reason to ensure that the criminal offender is brought to trial and punished to restore the peace—a good of considerable moral significance. To the extent that tracebacks can reliably be used to identify an attacker, they function to secure the important moral good of restoring the public peace by bringing wrongdoers to justice. Accordingly, responses motivated by such an objective are intended to secure an important moral value.[14]

Moreover, it also seems clear that such goods are important enough to justify comparatively minor infringements of the property rights of innocent persons. If the only way that a private security officer can apprehend a robbery suspect is to commit a trespass against the property of an innocent person, it seems clear that she is justified in doing so under the Necessity Principle. The moral value of bringing the offender to justice and thereby restoring the public peace greatly outweighs the moral disvalue of a simple trespass onto the land of an innocent party.

The problem with benign responses, however, is that it will frequently be unclear whether they are reasonably calculated to succeed in identifying culpable parties. As noted above, any reasonably sophisticated hacker will attempt to conceal her identity by staging the attack from innocent agent machines. Indeed, it is possible for a sophisticated attacker to further insulate herself from discovery by compromising one set of innocent machines to control another set of innocent machines that will be used to stage the attack—a process that can be iterated several times. In such cases, the attacker will interpose several layers of innocent machines between herself and the victim. But the greater the number of layers between attacker and victim, the less likely benign responses will succeed in identifying the culpable party. Although benign response can be highly effective in identifying the culpable parties in attacks staged directly from the hacker's machine, the probability of success drops dramatically with each layer of machines between attacker and victim. Indeed, it is fair to say that the likelihood of identifying the culpable parties in sophisticated attacks by benign responses is morally negligible.

This seems to imply that the victim of a digital attack cannot permissibly adopt benign responses under the Necessity Principle. Unless she has some special reason to think that the attack is being staged directly from the hacker's own machines without the use of benign agent machines or networks, she will not have adequate reason to think that benign measures will succeed in identifying the culpable parties and will not

[14]Not all benign responses are motivated by a desire to prosecute the wrongdoer. Many firms would prefer to avoid prosecution to avoid the unfavorable publicity that might result from the disclosure of security breaches. The above reasoning would not justify benign responses in these cases.

be ethically justified, under the Evidentiary Principle, in acting upon the Necessity Principle.

Nevertheless, it is important to emphasize that the analysis here is limited to current traceback technologies with their limitations. Many researchers are making considerable progress in improving the reliability and efficacy of traceback technologies.[15] Indeed, one might reasonably expect researchers to eventually improve these technologies to the point where they are sufficiently efficacious in identifying culpable parties that they can generally be justified under the Necessity Principle as bringing about the greater moral good of identifying culpable parties to an attack.

8.4.5 The Relevance of Consent

The preceding analysis presupposes that the victim of a digital attack does not have express or implied permission to causally impact the machines of innocent owners. One might think that owners of agent machines somehow consent to being affected by active response measures. If owners of affected machines have consented to such effects, then they have waived any general moral immunity from active response.

Clearly, there is no general reason to think that owners of agent machines have explicitly or expressly consented to either having their machines used for an attack or being targeted by aggressive countermeasures. In the absence of any other reason to think aggressive countermeasures against these machines are permissible, victims would be committing a wrong against the owners under the Evidentiary Principle should they direct aggressive countermeasures at these machines.

In some rare instances, persons can be presumed to have "tacitly" or "impliedly" waived a right on the basis of some nonexpressive behavior not intended to effect a waiver. Indeed, in some instances, a person's failure to object to some act can be treated as tacit consent to that act. For example, there is little disagreement about the justice of the legal rule that treats an attorney's failure to object to something opposing counsel has done as having waived the objection.

Accordingly, one might argue it is reasonable to infer that owners of agent machines used in a digital attack have consented to being targeted by active response. On this line of analysis, the failure of such owners to protect against unauthorized entry with a firewall is reasonably construed as consent to entry in cases where it is needed to investigate or defend against a digital attack staged from their machines. On this line of reasoning, someone who fails to take such precautions is reasonably thought to be sufficiently indifferent about the prospects of intrusions that she may be presumed to consent to them.

There are a couple of problems with this line of reasoning. First, it would not only imply consent to the victim's intrusion, but would also imply consent to the attacker's intrusion—a result that is sufficiently implausible to warrant rejecting any claims that imply it. Second, failure to implement a firewall is no more reasonably construed as consent to entry than failure to lock the door to one's car is reasonably construed as consent to enter one's car. One might forget to take such precautions for any number of

[15]See http://footfall.csc.ncsu.edu, which documents some intriguing advancements in these technologies.

reasons without being indifferent about unwanted entries. Moreover, in the case of computer intrusions, one might simply not know about the available security options.

A somewhat more plausible argument for treating failure to take adequate security precautions is grounded in ethical principles that impute a duty of reasonable care to protect others from foreseeable harm. In cases where one person's negligence potentially puts another person at risk, considerations of fairness require imputing some responsibility or disadvantage to the former person that must ordinarily be voluntarily accepted. If, for example, you negligently disclose my whereabouts to someone who wants culpably to harm me, you might thereby obligate yourself to do something for me that you ordinarily would not be obligated to do—perhaps hide me in your home.

One might, then, argue that persons who fail to take reasonable precautions to prevent unwanted computer intrusions and whose computers are used to stage an attack have tacitly consented to aggressive and benign active defense measures directed at their computers by the victims of those attacks. Since their negligence has wrongfully put innocent persons at risk, they have released the victims of the attacks from any duties they otherwise might have had to refrain from benign or aggressive active defense.

If ordinary intuitions and practices are correct, this line of reasoning will not justify directing aggressive measures at owners of compromised machines. Ethical principles of negligence are not generally thought to justify aggression against negligent parties; they would not, for example, justify me in attacking a person who has negligently injured me or damaging her property. Rather they are thought to require a person to compensate parties for injuries proximately caused by her failure to take reasonable precautions to protect such parties from injury; this, of course, is how such principles are interpreted and applied by the courts under tort law.

Whether ethical principles regarding negligence might justify benign responses is a much more difficult issue that cannot be addressed here. Admittedly, there is little in ordinary practices that would justify an inference that these principles allow benign responses. There are simply no obvious analogs in ordinary practices to digital attacks staged from innocent machines.

Even so, the idea that the owner of a compromised agent machine might have waived any immunity she would have otherwise had to benign responses is not obviously unreasonable. If, for example, it is reasonable to think that such persons have a duty, at the very least, to contribute to compensating the victim of a digital attack for injuries sustained during the attack, it also seems reasonable to think that such persons have a duty to permit victims to commit an intrusion for the purpose of tracing the attack back to its source. If, during the course of an attack, the victim had adequate reason to think that (1) benign responses would successfully identify the attacking party and (2) owners of compromised machines negligently failed to take reasonable precautions to prevent their machines from being used in a digital attack, she might very well be justified in directing benign responses against those machines.

For all practical purposes, however, the argument is moot. Since victims will rarely be able to gather, during the course of a digital attack, adequate evidence for thinking either that benign responses would be successful or that owners of compromised machines have failed to take reasonable precautions, they will not be able to justify

adopting these responses under ethical principles of negligence. For this reason, the Evidentiary Principle seems to preclude adopting benign responses on the strength of ethical principles of negligence—assuming, of course, that these principles are even applicable.

8.4.6 The Inadequacy of Law Enforcement Efforts

There is one last argument that can be made in defense of the idea that it is permissible for private individuals to undertake various active defense measures.[16] The argument rests on the idea that the state may legitimately prohibit recourse to self-help measures in dealing with a class of wrongful intrusions or attacks only insofar as the state is providing minimally adequate protection against such attacks. If (1) digital intrusions are resulting in significant harm or injury of a kind that the state ought to protect against and (2) the state's protective efforts are inadequate, then private individuals, on this line of reasoning, are entitled to adopt active defense measures that conduce to their own protection.

Both antecedent clauses appear to be satisfied. Depending on the target and sophistication of the attack, an unauthorized digital intrusion can result in significant financial losses to companies. For example, an extended distributed denial of service attack that effectively takes Amazon.com offline for several hours might result in hundreds of thousands of dollars of business going to one of its online rivals. In the worst-case scenario, these financial losses can result in loss of value to shareholders and ultimately loss of jobs. It seems clear that the harms potentially resulting from digital intrusions fall within a class that the state ought to protect against.

Further, there is good reason to think that the state's protective efforts are inadequate. At this point in time, law enforcement agencies lack adequate resources to pursue investigations in the vast majority of computer intrusions. But even when resources allow investigation, the response might come after the damage is done. Law enforcement simply has not been able to keep pace with the rapidly growing problems posed by digital attackers.

There are a variety of reasons for this. Most obviously, the availability of resources for combating cybercrime is constrained by political realities: if the public is vehemently opposed to tax increases that would increase the resources for investigating cybercrime, then those resources will not keep pace with an increasing rate of intrusions. But, equally importantly, there are special complexities involved in investigating and prosecuting digital intrusions. First, according to Mitchell and Banker, investigation of digital intrusions is resource-intensive: "whereas a typical (non-high-tech) state or local law enforcement officer may carry between forty and fifty cases at a time, a high-tech investigator has a full time handling three or four cases a month." Second, most sophisticated attacks will pose jurisdictional complexities that increase the expense of law enforcement efforts because such attacks will frequently involve crossing jurisdictional lines. For example, an attacker in one country might compromise machines in another country in order to stage an attack on a network in yet a third country.

[16]See Mitchell and Banker (1998).

Although such considerations show that the growing problem associated with digital intrusions demands an effective response of some kind, they fall well short of showing that it is permissible, as a general matter, for private parties to undertake benign or aggressive active defense measures. The underlying assumption is that private individuals can adequately do what the state cannot—namely, protect themselves adequately from the threats posed by digital intrusion.

At this time, however, there is very little reason to think that this underlying assumption is correct. For starters, invasive benign measures intended to collect information are likely to succeed in identifying culpable parties only in direct attacks staged from the attacker's own computer; such measures are not likely to succeed in identifying parties culpable for intrusions that are staged from innocent machines. Since an attacker sophisticated enough to stage an attack likely to result in significant damage is also likely to be sophisticated enough to interpose at least one layer of innocent machines between her and her target, there is little reason to think that invasive investigatory measures are likely to achieve their objectives in precisely those attacks that are likely to result in the sort of damage that the state is obligated to protect against.

Moreover, aggressive measures are not likely to conduce to the protection of the victim in any reasonably sophisticated attack. As noted above, aggressive countermeasures are not usually calculated to result in the cessation of the attack and can frequently result in escalating the attack; for this reason, such countermeasures are not likely to succeed in purely defensive objectives. Unfortunately, they cannot succeed in achieving legitimate punitive objectives in attacks staged from innocent machines. Punitive measures directed at the innocent agents do nothing by way of either punishing the ultimate source of the attack or deterring future attacks. A reasonably sophisticated attacker who knows her target will respond with aggressively punitive measures will simply evade the effects of those measures by interposing an additional layer of innocent machines between her and her target.

REFERENCES

Auty, C. (2004). Political hacktivism: tool of the underdog or scourge of cyberspace? *ASLIB Proceedings: New Information Perspectives*, 56, 212–221.

Huschle, B. (2002), Cyber disobedience: when is hacktivism civil disobedience? *International Journal of Applied Philosophy*, 16(1), 69–84.

Himma, K.E. (2006a). Hacking as politically motivated digital civil disobedience: is hacktivism morally justified? In: Kenneth, E.H. (Ed.), *Readings on Internet Security: Hacking, Counterhacking, and Other Moral Issues*. Jones & Bartlett, Sudbury, MA.

Himma, K.E. (2006b). The ethics of active defense. In: Kenneth, E.H. (Ed.), *Readings on Internet Security: Hacking, Counterhacking, and Other Moral Issues*. Jones & Bartlett, Sudbury, MA.

Himma, K.E. and Dittrich, D. (2006c). Hackers, crackers, and computer criminals. *The Handbook of Information Security.* John Wiley & Sons.

Levy, S. (1984). *Hackers: Computer Heroes of the Computer Revolution*. Delta Trade Paperbacks, New York.

Manion, M. and Goodrum, A. (2000). Terrorism or civil disobedience: toward a hacktivist ethic. *Computers and Society*, June, 14–19.

Mitchell, S.D. and Banker, E.A. (1998). Private intrusion response. *Harvard Journal of Law and Technology*, 11(3), 710.

Spafford, E. (1992). Are computer hacker break-ins ethical? *Journal of Systems Software,* 17(1), 41–48.

Tavani, H. (2007). The conceptual and moral landscape of computer security. In: Himma, K.E (Ed.), *Internet Security*: *Hacking*, *Counterhacking*, *and Society.* Jones & Bartlett, Sudbury, MA.