IJIE
International Journal of Information Ethics Vol. 2 (11/2004)

Kenneth Einar Himma

The Ethics of Tracing Hacker Attacks through the Machines of Innocent Persons

Abstract:

Victims of hacker attacks are increasingly responding with a variety of "active defense" measures, including "invasive tracebacks" that are intended to identify the parties responsible for the attack by tracing its path back to its original source. The use of invasive tracebacks raise ethical issues because, in most cases, they involve trespassing upon the machines of innocent owners. Sophisticated hackers attempt to conceal their identities by routing their attacks through layers of innocent agent machines and networks that are compromised without the knowledge of the owners. The use of invasive traceback technologies in such cases, then, involves an act is presumptively problematic from an ethical standpoint: intentionally entering upon the property of an innocent person without her consent constitutes a prima facie trespass.

I argue that there is no ethical principle that would justify the use of invasive tracebacks by private persons or entities (as opposed to governmental persons or entities). First, I argue that invasive tracebacks cannot be justified under the Defense Principle, which allows one person to use proportional force to defend herself or other innocent persons from attacks. Second, I argue that, in ordinary cases, the use of an invasive traceback impacting innocent persons cannot be justified under the Necessity Principle, which permits the infringement of an innocent person's rights when necessary to achieve a significantly greater good. Since these are the only applicable principles, I conclude that, in the absence of special circumstances, it is not ethically permissible for private parties and entities to implement invasive traceback technologies.

Agenda

Author:

Prof. Dr. Kenneth Einar Himma:
- Seattle Pacific University, Department of Philosophy, 3307 Third Avenue West, Seattle, WA 98119, USA
- ✉ himma@spu.edu

ISSN 1614-1687

Introduction

Hackers are posing an ever-greater threat to Web-based governmental and corporate activities. While hackers are increasing in both number and sophistication, the resources available to law enforcement agencies is increasing, if at all, at a much slower rate. Hackers are clearly winning the battle with law-enforcement agencies, which must content themselves with investigating and prosecuting only the most spectacular cases.

Not surprisingly, private firms have begun to take matters into their own hands, responding to hacker attacks with a variety of "active defense" measures.[i] Some of these responses are aggressive in the sense that they are intended to inflict the same kind of harm on the attacker's machine or network as the attack is intended to have on the victim's machine or network. While private firms may sometimes employ these measures for purely defensive reasons, they are also frequently motivated by a desire to retaliate and deter future attacks: in many cases, the attack can be stopped with far less aggressive measures.[ii]

The use of aggressive measures by private firms is ethically problematic for a variety of reasons. To begin, most sophisticated attacks are staged from a layer of machines that have been compromised without knowledge or fault on the part of their owners; in such cases, aggressive active defense deliberately causes harm to innocent persons – something that is, at the very least, presumptively problematic. Moreover, in sophisticated attacks, aggressive measures are more likely to escalate hostilities than to end them. Finally, it is generally accepted that it is the province of the state, and not the aggrieved individual, to direct force at an offender for the purposes of punishing and deterring wrongdoing; for this reason, aggressive active defense is not unreasonably characterized as wrongful "vigilantism."

A different (and more difficult) set of ethical issues, however, arises in connection with less aggressive active defense measures that attempt to identify the parties responsible for a digital attack by tracing the path of the attack back to its original source. There are a variety of "traceback" technologies and techniques available to victims of Internet-based attack. The most benign of these techniques is simply to take attacking IP addresses – information contained in the attacking traffic itself – and then conduct a "whois" lookup for that address at the various domain registry services. In contrast, the more invasive techniques and technologies attempt to identify the identity of parties culpable for a digital attack by *entering into* compromised machines or networks.

While there is little reason to think that the more benign technologies are unethical, the use of "invasive tracebacks" raises ethical issues when hacker attacks are staged from innocent agent machines. Although the use of invasive tracebacks does not cause harm to these agents, it involves unauthorized entry upon the property of innocent persons – something that is presumptively wrong: intentionally entering upon the property of an innocent person without her consent constitutes a *prima facie* trespass. Accordingly, the use of such technologies can be justified only insofar as it falls within the application-conditions of some generally accepted moral principle that protects a more important interest than the interest in being free from trespass.

In this essay, I argue that there is no ethical principle that would currently justify the use of invasive tracebacks by *private* persons or entities (as opposed to governmental persons or entities).[iii] To begin, I argue that invasive tracebacks cannot be justified under the Defense Principle, which allows one person to use proportional force to defend herself or other innocent persons from attacks. The problem is that tracebacks are used to identify parties and cannot, strictly speaking, be used to "defend" against an attack.

Further, I argue that, in ordinary cases, the use of an invasive traceback impacting innocent persons cannot be justified under the Necessity Principle, which permits the infringement of an innocent person's rights when necessary to achieve a significantly greater good. The problem here arises because the use of tracebacks can result in a variety of significant intra- and inter-cultural harms that are not balanced by a sufficiently greater moral good because tracebacks are currently unreliable in identifying the parties responsible for an attack. Since these are the only applicable principles, I conclude that, in the absence of special circumstances, it is not ethically permissible for private parties and entities to implement invasive traceback technologies.

Two preliminary observations are in order here.[iv] First, the arguments in this essay apply only to existing traceback technologies. It is not unreasonable to think that traceback technologies will continue to improve over time as researchers

develop better techniques and cleaner codes. Thus, we can reasonably expect that future traceback technologies will not have the same morally significant limitations of existing technologies; they will likely be more efficacious with fewer unintended harmful inter- and intra-cultural consequences. If so, then future technologies might very well be justified under the Necessity Principle.

Second, the analysis here is not grounded in any general ethical theory like consequentialism, the ethic of care, or Kantianism constructivism. Rather, as is common in applied ethics, the analysis is grounded in principles and case-judgments which figure prominently in ordinary ethical practices. Accordingly, the analysis begins by identifying ethical principles that I think that most people would accept as correct and proceeds by attempting to identify the implications of those principles.

This means that the analysis here is capable of persuading only those persons who accept the principles and case-judgments that ground it. While these principles and judgments are incorporated into the law of every Western industrialized nation and hence widely accepted as just, they might not be universally accepted in all cultures. If not, then the analysis here will not persuade persons in all cultures – though I would be surprised if something like these principles were not universal.

Innocent Persons and the Defense Principle

At the outset, it is important to realize that the risk that active defense measures will impact innocent machines is not just "theoretical."[v] *Most* sophisticated attackers attempt to conceal their identities by compromising innocent machines and staging their attacks from these "agents" – which are frequently located all over the world. To adequately defend against or investigate an attack, active countermeasures will have to be directed, at least initially, at the agents used to stage the attack. Accordingly, it is nearly inevitable that any reasonably efficacious active defense strategy will impact innocent persons.

Anyone sophisticated enough to implement an active defense strategy even remotely likely to succeed in countering an attack presumably realizes this. Indeed, one could not make an *informed* choice of active defense strategies without understanding the structure of the attack and the various countermeasures most likely to stop it. And anyone who understands these things must surely know that an efficacious response will likely impact innocent machines in a variety of ways that are potentially problematic from the standpoint of morality.

While it is generally impermissible for one person to infringe the rights of innocent persons, there are exceptions.[vi] One obvious example is the principle that allows us to use proportional force when necessary to defend against an attack:

> **The Defense Principle:** It is ethically permissible for one person to use force to defend oneself or other innocent persons against an attack provided that (1) such force is proportional to the force used in the attack; (2) such force is necessary either to repel the attack or to prevent the attack from resulting in harm of some kind; and (3) such force is directed at, and likely to harm, only those persons who are responsible for the attack.While there is considerable disagreement among cultures about the content of moral principles, most cultures accept something like the Defense Principle, which is also incorporated into the criminal law of nearly every developed legal system in the world.

While there is considerable disagreement among cultures about the content of moral principles, most cultures accept something like the Defense Principle, which is also incorporated into the criminal law of nearly every developed legal system in the world.

The Defense Principle is generally thought to allow force against innocent persons in one fairly narrow situation. While I may never direct force against innocent *bystanders* to defend against an attack, I may direct force against what are plausibly characterized as innocent *attackers*. If, for example, I am attacked by someone who is obviously insane and not morally responsible for his actions, I may, under the Defense Principle, defend myself against him with proportional force. Despite the fact that the attacker is innocent of any wrongdoing because incapable of instantiating a culpable mental state, I may direct force against him under the Defense Principle as long as it is necessary to defend against the attack. This interpretation of the Defense Principle is nearly unquestioned among theorists and laypersons.[vii]

Though innocent agent machines seem to fall within the application-conditions of the Defense Principle as innocent attackers, this principle cannot justify

the use of invasive tracebacks for a couple of reasons. First, the use of invasive tracebacks does not necessarily involve anything that is plausibly characterized as force. It is part of the conceptual nature of force that it be capable of inflicting damage, injury, or harm. It makes sense to characterize redirecting a DoS attack back at the attacker as *force* because overloading a network results in something that is fairly characterized as *harm*; if the victim's business is taken offline, she will lose business – something that clearly involves an injury of sorts. But while invasive tracebacks involve entering the machines of other persons, such acts do not necessarily inflict damage, injury, or harm. Insofar as these traceback technologies do not involve anything that necessarily inflicts (or attempts to inflict) damage, injury, or harm, they are not properly characterized as "forceful" and hence cannot be justified by the Defense Principle.

Second, and more importantly, invasive tracebacks do not have any features that would either *repel* the attack or *prevent* the attack from resulting in harm to the victim. The point of using a traceback technology is to identify the culpable attacker by following an ongoing attack back through intermediate sources to its origin. Indeed, insofar as such technologies do not involve anything plausibly characterized as force capable of inflicting an injury, they *could not* repel an attack. Further, insofar as such technologies do not involve anything that enables the victim to escape from the act, they do nothing do prevent any harm; the only ways to prevent an attack from resulting in harm is to either repel the attack or escape.

At this juncture, invasive tracebacks can succeed in identifying the culpable parties only while the attack is ongoing. In this sense, they resemble technologies for tracing a telephone call; a telephone call can be traced only while the calling party remains on the line. It is no accident, then, that invasive tracebacks do not incorporate techniques plausibly characterized as forceful; the concomitant use of force would diminish the likelihood of identifying the parties by increasing the probability that the attacker will end the attack. These technologies can succeed only insofar as they do nothing that would repel or defend against the attack. Accordingly, since the Defense Principle can justify only the use of measures intended to repel an attack or prevent harm, the use of invasive traceback technologies cannot be justified by reference to the Defense Principle.

Innocent Persons and the Necessity Principle

There is one other widely-accepted ethical principle that allows one person to infringe the rights of innocent persons that might justify the use of invasive traceback technologies. An example will help to develop the principle. Assume that the following are all true: (1) Attacker is attempting to set Victim's house on fire by throwing Molotov cocktails at Victim's house; (2) Victim's child is in the house; (3) Attacker is throwing these cocktails from the property of Innocent Bystander who is away on a business trip; and (4) the only way Victim can stop the attack before it succeeds is to trespass onto Bystander's land. Most people (indeed, in most cultures) would agree that, under these circumstances, it is permissible for Victim to trespass onto Bystander's property. Though such an act clearly *infringes* Bystander's property rights, it does not *violate* those property rights precisely because it is morally justified.[viii]

There are four considerations that explain this judgment. First, Victim will achieve great moral value by saving her child's life and her dwelling from a culpable attack. Second, Victim cannot achieve such moral value without trespassing onto Bystander's land. Third, the threat to Victim's interests is much greater, morally speaking, than the threat to Bystander's interests. If Attacker succeeds, then an innocent child will be killed and Victim will be forcibly dispossessed of her dwelling without any claim of right. The threat to Bystander's interests involves no more than a temporary presence on her land since Victim does not need to cause any damage to the land in order to stop Attacker's assault and thereby save her home and child. Finally, Victim's objective is a morally respectable one – namely, to save her child's life and home from a culpable attack.

Putting these four features together suggests an uncontroversial general principle that limits the moral immunity of innocent persons to measures that potentially infringe their rights:

> **The Necessity Principle:** It is ethically permissible for one person *A* to infringe a right ρ of an innocent person *B* if and only if (1) *A*'s infringing of ρ is reasonably likely to result in great moral value; (2) the good that is protected by ρ is significantly less valuable, morally speaking, than the good that *A* can bring about by infringing ρ; (3) there is no other way for *A* to bring about

this great moral value that does not involve infringing *ρ*; and (4) *A*'s attitude towards *B*'s rights is otherwise properly respectful.[ix]

While this formulation is somewhat more technical than is customary, something like this principle is widely accepted across cultures and, like the Defense Principle, incorporated into the criminal law of nearly every developed legal system.

It is worth noting that the Necessity Principle augments the Defense Principle by allowing some action 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 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.[x] But insofar as the Necessity Principle requires the achievement of a *significantly* greater good, it will not allow a person to direct force at an innocent bystander that is proportional to the force of the attack.

Though there is some overlap between the two principles, the rationales for the two principles are clearly different. On the most common conception of the right to self-defense, the culpable behavior of the attacker "forfeits" her right not to be attacked – at least to the extent that proportional force is involved; someone who threatens your right to life by shooting at you has forfeited her right to life for the duration of the attempt on your life. But it is clear that this cannot be what explains the validity of the Necessity Principle since one can forfeit a right only by expressly consenting to its forfeiture or by committing an act that directly infringes the rights of innocent persons. Since, by definition, an innocent bystander has not committed a culpable act and since we have no reason to think that she consents to the forfeiture of any right, the considerations that explain the right of self-defense cannot explain the Necessity Principle.

The most plausible remaining explanation is that the scope of many rights simply does not extend to situations in which a significantly greater good can be achieved only by infringing the relevant interest. On this line of analysis, my property right to exclude persons from using or being on my land does not extend to situations in which a person can save a life from culpable attack only by entering onto my land without my permission. In such cases, the person defending against such an attack has a moral permission/liberty to enter onto my land as long as she otherwise evinces proper respect for my interests and rights.

Applying the Necessity Principle to Invasive Tracebacks

An Epistemic Precondition for Justifying Action under an Ethical Principle

Before evaluating the application of the Necessity Principle to invasive tracebacks, I should note that there is an evidentiary (or epistemic) precondition that must be satisfied in order to justifiably take action under any ethical principle: one can be morally justified in taking action under an ethical principle only to the extent that one has adequate reason to believe that its application-conditions are satisfied. To see this, consider that Paul Hill argued that he was justified in murdering John Bayard Britton, an abortion provider, by the Defense Principle, which allows deadly force in defense of the lives of innocent moral persons against culpable attack.[xi] Since, according to Hill, fetuses are moral persons from conception and since murdering Britton was necessary to save the lives of fetuses he would culpably abort, he was justified in killing Britton under the Defense Principle – just as he would be justified under that principle in killing someone who was trying to murder a newborn infant.[xii]

Nevertheless, Hill's murder is not justified under the relevant the Defense Principle precisely because the epistemic preconditions for its application were not clearly satisfied. Insofar as reasonable persons disagree sharply on whether fetuses are moral persons from the moment of conception, much more argument is needed to provide adequate reason to believe this is the case. Since Hill lacked morally adequate reason to believe that the principle allowing deadly force in defense of innocent persons applied to *fetuses*, he could not be justified under the Defense Principle in killing Britton and was rightly convicted of murder. As a general matter, a person who takes forceful action against a person without adequate reason to think some moral principle's application-conditions are satisfied commits a moral wrong against that person.

It follows that the victim of an Internet-based attack can justifiably take action under the Necessity Principle only if she has adequate grounds for believing that its application-conditions are satisfied. The Necessity Principle permits an agent to perform act *a* knowing that it will infringe an innocent person's rights if and only if three conditions are satisfied: (1) the good secured by *a* significantly outweighs the evil that is done; (2) there is no other

way to achieve the significantly greater good than to do *a*; and (3) the performance of *a* is reasonably likely to succeed in achieving the significantly greater good. Accordingly, the victim of an Internet-based attack can justifiably take action under the Necessity Principle only if she has adequate grounds for believing 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* is reasonably likely to succeed in achieving the greater moral good. If the victim of such an attack of any kind lacks adequate evidence for any of these three propositions, she cannot justifiably act under the Necessity Principle. If she nonetheless acts in a way that infringes an innocent person's rights and if there is no other moral principle that would justify doing so, she has committed a moral wrong against that person. It is argued below that, absent special circumstances, only two of the three conditions above are satisfied with respect to the private use of invasive tracebacks.

Identifying and Weighing the Relevant Goods and Evils

In evaluating the permissibility of invasive tracebacks under the Necessity Principle, we can rule out one important good at the outset. Since these tracebacks are not designed to repel attacks or prevent the harms that result from such attacks, they cannot achieve the significant moral good of minimizing the victim's losses or damages. While this is a good that *defensive* measures are capable – at least in principle – of securing, invasive tracebacks are not, strictly speaking, defensive measures in the relevant sense. Accordingly, they are not – and *cannot be* – used to prevent the significant economic losses that frequently result from, say, DDoS attacks on commercial websites.

Even so, we need not look far for an important moral good that invasive tracebacks are contrived to secure. Criminal attacks are traditionally regarded as offenses against the general public – and not just against the individual victim or victims – for a couple of reasons. First, criminal attacks directed at one member of the community can, and frequently do, have harmful effects on other members of the community. When, for example, a defendant commits a murder, it can cause considerable fear and anxiety that can lead other persons in the community to modify their behavior in morally significant ways. Second, and equally importantly, criminal attacks always violate the legitimate expectations of the public and thereby breach the peace against the public.

Accordingly, though the individual victim has a special interest in wanting to see the criminal offender brought to trial and punished, the public also has a compelling interest in the fate of the criminal offender. Legitimate punishment of the guilty not only gives the offender what, as a moral matter, she deserves, but also helps to restore the peace. As long as the offender remains at large, the community is likely to continue to experience the sort of anxiety that can have a significant chilling effect on the exercise of their liberties. Bringing the offender to justice restores the peace by alleviating such collective anxiety and vindicating the legitimate expectations of the community.[xiii] Additionally, public punishment of the offender serves as a deterrent to future attacks and thereby helps to reduce the probability of further breaches of the peace. It is utterly uncontroversial that the restoration of the peace following a criminal offense is a good of considerable moral significance.

To the extent that invasive tracebacks can reliably be used to identify the culpable source of an Internet-based attack, they function to secure the important moral good of restoring the public peace by bringing a wrongdoer to justice. Identifying the party responsible for an Internet-based attack enables the state to bring that party to justice, to alleviate the public anxieties that typically follow criminal behavior, and to deter future would-be hackers. In theory at least, then, the use of invasive tracebacks conduces to moral goods of tremendous importance.

It is also uncontroversial that the magnitude of such goods is sufficient to justify comparatively minor infringements of an innocent person's rights if necessary to restore the peace. Suppose, for example, that an offender who has committed a robbery is attempting to escape from a private security officer who is chasing her down a public street. The robber's path eventually takes her onto the land of a private citizen who is away from her home at the time. If the only way that the security officer can apprehend the shoplifter is to come uninvited upon the innocent party's land and commit what would otherwise be a trespass, then it is clear, under the Necessity Principle, that it is morally permissible for her to do so.[xiv] The moral value of restoring the public peace greatly outweighs the moral disvalue of a simple trespass onto the land of an innocent party.

Likewise, the moral value of restoring the public peace greatly outweighs the moral disvalue of a simple digital trespass onto an innocent party's computer or network. As long as the user of such technologies does not infringe other rights of the innocent parties (by, for example, examining or troying files obviously unrelated to the attack), the relevant moral benefits associated with restoring the public peace greatly outweigh the relevant moral costs. Insofar as invasive tracebacks are used only to gather evidence that enables prosecutors to bring the culpable parties to justice, their use conduces to a significantly greater moral good.

Are there Other Plausible Methods for Identifying Culpable Parties?

At this point in time, it is reasonable to think that there are no other methods for identifying the culpable parties to an Internet-based attack that are generally reliable. Even if we assume that public law-enforcement agencies have some special ability to identify attackers during the course of an attack, they are typically slow to respond; as the point has been recently put, "Unless your company is a large organization[,] whatever help is forthcoming from agencies like the FBI will take a relatively long time especially in 'Internet time.'"[xv] Since the probability of identifying a culpable attacker is highest during the attack,[xvi] the inability of public law-enforcement agencies to respond in a timely way significantly diminishes the likelihood of identifying the ultimate source of an attack.

This should not be construed as a criticism of law-enforcement agencies in any particular culture. These agencies have to do the best they can with whatever resources the taxpaying public is willing to subsidize. Given that digital attacks are non-violent crimes against property and that resources are extremely limited, it is perfectly appropriate for law-enforcement agencies to treat them with less urgency than violent crimes against persons or property. If there is any fault here, it ultimately lies with the legislatures that fail to adequately fund law enforcement agencies.

But the absence of a consistently timely response in such cases does suggest, as a general matter, that the likelihood that law-enforcement agencies will be able to determine the identity of culpable attackers is comparatively low. If a timely response is needed to maximize the probability of identifying culpable parties, then it follows that an untimely response diminishes the probability of doing so. Accordingly, since tracebacks can be implemented by the victims of an attack more quickly than by any other party, the use of traceback technology by the victims arguably provides the *only* genuine opportunity to identify the ultimate source of an attack – information that law-enforcement agencies *must* have in order to bring about the great moral good associated with restoration of the public peace.

Again, the Necessity Principle will not justify any other infringements of the rights of innocent persons than are necessary to restore the public peace. Someone who, for example, attempts to gain access to content on an innocent machine not needed to identify the culpable attacker commits a violation of the owner's rights; such an infringement is not justified under the Necessity Principle because it is not necessary to achieve the greater moral good of restoring the peace. For these reasons, the Necessity Principle limits the private utilization of traceback technology to only those uses essential to gathering evidence that will conduce to bringing the culpable attacker to justice; any other use by private entities is morally problematic.

The Efficacy of Invasive Tracebacks in Identifying Culpable Parties

So far, two of the three conditions needed to justify the private use of invasive tracebacks under the Necessity Principle seem to have been satisfied. First, it seems clear that the moral value involved in bringing wrongdoers to justice and thereby restoring the public peace significantly outweighs the moral disvalue of committing a simple digital trespass. Second, it seems equally clear that, at least in the absence of a timely response from law enforcement agencies, there is no other method for identifying culpable parties that is reasonably likely to succeed. Whether the private use of invasive tracebacks can be justified under the Necessity Principle, then, turns on whether the third condition is satisfied – that is, whether there is adequate evidence that invasive tracebacks are reasonably likely to succeed in identifying culpable parties.

In thinking about this third condition, it is crucial to reiterate that any reasonably sophisticated hacker will attempt to put some distance between her and her victim by attacking the victim through third-party intermediary machines. A sophisticated hacker will usually compromise a set of vulnerable agent machines or networks in such a way as to permit her to control those machines from another remote machine (e.g., her home machine), thereby interposing a layer of insulation (or a "hop in the chain") between her and her victim: the immediate

source of the attack is the set of agent machines controlled by the hacker's remote machine, which is the ultimate source. And hackers are not limited to one layer of insulation: it is possible to compromise two sets of intermediate machines, using one to stage an attack directly from the other. In such cases, the attacker interposes two hops in the digital-causal chain that links the attacker's machine with the victim's machine.

The efficacy of any particular traceback technology, including invasive technologies, in identifying culpable parties depends on the structure of the attack and, in particular, on the causal proximity of the culpable party's machine to the victim's machine. The greater the number of hops in the causal chain linking attacker and victim, the less likely that any traceback technology will succeed in identifying the ultimate source of an attack. While tracebacks can be highly effective in tracing attacks that are staged directly from the hacker's machine, they are considerably less effective in tracing attacks that are routed through layers of intermediate agent machines or networks – and the probability of success drops dramatically as the hacker adds additional hops in the chain. If the hacker is reasonably careful in selecting mechanisms for controlling the different layers of machines, the probability that the culpable party can be identified by tracebacks is fairly characterized as negligible.

As an empirical matter, direct attacks are becoming less common as hackers become more sophisticated. As one prominent security expert explains:

> *"[A]ttacker[s] sitting at home on their PCs very rarely (unless they are rather naïve) will connect to a PPP server (or use a broadband/DSL direct IP connection) and then attack some site. This is just too easy to trace back. Instead, they will use one or more (the more, the better) compromised systems...."*[xvii]

At this point in time, then, only the most naïve hackers would stage direct attacks from their own machines or networks. Any reasonably sophisticated hacker will attempt to insert as many hops in the chain between her and her victim as is needed to minimize the likelihood of being identified.

But this means that the victim of an Internet-based attack will be justified in using invasive tracebacks under the Necessity Principle only insofar as she has adequate reason to think that the attack is being staged directly from the hacker's own machines without the use of intermediate agent machines or networks. As will be recalled, the third condition for justifiably using invasive tracebacks under the Necessity Principle is that there must be adequate reason to believe that such a measure is reasonably likely to succeed in bringing about the greater moral good of identifying the culpable parties. Since it is uncontroversial that invasive tracebacks are reasonably likely to succeed only in direct attacks, the third condition will not be satisfied unless the victim has minimally adequate evidence for believing that the attack is direct.

While it is undoubtedly true that there may always be cases in which this is true, these cases are, at this point in time, the exception and not the rule – and will become increasingly rare as hackers generally become more sophisticated not only with respect to the techniques they adopt but also with respect to how they convey those techniques to other would-be hackers. Absent special circumstances or special knowledge on the part of the victim contemplating the use of invasive tracebacks, the presumption should be that the use of invasive tracebacks is not likely to succeed in identifying the culpable attackers. For this reason, the moral disvalue associated with trespassing against the innocent agent machines cannot be justified, in ordinary cases, under the Necessity Principle by the significantly greater moral value of bringing the wrongdoer to justice and thereby restoring the public peace.

Here it is important to emphasize again that the reasoning above applies only to existing technologies. One can reasonably expect that, as traceback technologies are improved, they will become increasingly efficacious in identifying culpable parties. Indeed, it is not inconceivable that they might very well be improved to such an extent that invasive tracebacks become so highly reliable in identifying culpable parties that a victim is justified in presuming in any given instance that executing a traceback will be successful in identifying culpable parties. This, of courses, does not, by itself, imply that using tracebacks is permissible because there might be problems that counterbalance such advantages. But it does imply that the reasoning in the preceding paragraph referring to existing technologies would not apply to sufficiently efficacious technologies.

But insofar as current technologies are comparatively unreliable in identifying culpable parties, their use cannot be justified under the Necessity Principle as needed to bring about the

greater moral value of identifying culpable parties for the purpose of bringing them to justice.

Potential Impacts of Widespread Use on Intra- and Inter-cultural Community Building

While I think the foregoing analysis is sufficient to rule out the use by private parties of invasive tracebacks, there is an additional problem involved in trying to justify using invasive tracebacks by reference to the Necessity Principle. Up to now, I have considered only the direct effects of invasive tracebacks on the interests of the parties immediately involved in a digital attack: the victim, the owners of innocent agent machines, and the hacker. So far, the argument has considered only these effects in calculating the moral value and disvalue that would be achieved by the use of invasive tracebacks.

Unfortunately, the morally undesirable effects of any exchange between hacker, owners of innocent agent machines, and victim potentially extend far beyond just their interests. How such attacks are handled can have grave effects on those trust relationships within a particular culture that are essential to community-building efforts.[xviii] Consider, for example, a hacker who compromises the networks of a number of large U.S. businesses to stage attacks on the websites of *other* large U.S. businesses; such an attack would appear to the victim businesses to have been staged by its local competitors and likely interpreted as an act of corporate espionage.

Attacks like this can obviously impact a variety of intra-cultural trust relationships in harmful ways. Most obviously, they impact the relationships of the relevant U.S. businesses in ways that make them less likely to cooperate in socially useful ways and, indeed, may have the effect of making them far more likely to engage in unethical practices like corporate espionage. Less obviously, these attacks are likely to impact consumer trust in U.S. businesses because these attacks call attention to the security vulnerabilities of E-commerce.

The economic effects of these impacts within a culture are potentially great. The importance of E-commerce to economic activity in the U.S. has increased to the point where billions of dollars are at stake. Damage to "horizontal" trust relationships between competing businesses and to "vertical" trust relationships between consumers and businesses can result in significant economic losses and ultimately in the loss of jobs. Contractual economic activity has always involved a leap of faith; one must have trust that the other party is behaving in good faith and will fully abide by contractual terms. But the new Web-based information technologies require a greater trust from consumers and businesses for a variety of reasons. Since, for example, Web transactions the transmission of data from one theoretically vulnerable network to another, consumers must trust that businesses are not only operating in good faith, but also are making adequate efforts to secure the transmission of such data.

Moreover, how victims *respond* to attacks can also have significant effects on intra-cultural trust relationships. Suppose each of the victims in the above example launches a counterstrike directed against the agents from which the hackers is staging the attack. Now the innocent agent networks in the U.S. are also being directly attacked, but these attacks are being staged by U.S. businesses. These counterattacks are likely to compound the economic damage caused by the original attacks by increasing the damage to the various trust relationships.

Indeed, a situation in which major U.S. businesses are launching digital attacks against one another is fairly characterized as an intra-cultural "worst-case scenario." Consider John Pescatore's description of one possible scenario:

> *"My fear is that U.S. government agencies [involved in information warfare] will build in react capabilities. A smart hacker will launch a [denial-of-service] attack using those agencies' IP addresses and they all start attacking each other. The worst case is Amazon shoots eBay who shoots the IRS who shoots Cisco who shoots…."*[xix]

The idea that major U.S. corporations would engage in something that resembles cyberwarfare could have a variety of ramifying effects on socio-psychological and economic phenomena. Clearly, the intra-cultural impacts of aggressive countermeasures are potentially devastating.

Even the use of less aggressive active defense measures, like invasive tracebacks, is problematic from the standpoint of intra-cultural community-building. Imagine the likely reaction of the U.S. businesses in the example above to finding out that traceback technologies have been used to track the attack through *their* servers and networks. The same networks and servers from which a digital attack can be staged might also contain sensitive information about clients and customers. The

attempt by one U.S. company to trace a digital attack through the equipment of other U.S. companies can have significant effects not only on the relationships among the businesses, but also on the relationships between the businesses and their potential customers. The effects of adopting any aggressive or invasive active defense measure on intra-cultural community-building efforts can clearly result in profound moral disvalue.

The potential effects of such measures on *inter*-cultural community-building efforts are significantly more worrisome. Suppose, for example, that a hacker attack against commercial machines in the U.S. is staged from a number of compromised machines which include machines used by government officials in North Korea, a state that has made no secret of its attempt to develop a significant nuclear arsenal. The adoption of aggressive or invasive active defense measures by commercial firms against these machines has the potential to increase tensions between the U.S. government and the North Korean government, potentially putting millions of people at risk by derailing efforts to build community connections between two nations with nuclear weapons.

It is clear that the moral disvalue involved in the worst-case scenarios in both examples would outweigh the moral value to be achieved by the adoption of invasive tracebacks. In the worst-case scenario involving the intra-cultural example, the use of invasive tracebacks results in significant economic damage because it undermines the trust-relationships vital to cooperative economic activity even in a highly competitive economic environment like the U.S. In the worst-case scenario involving the inter-cultural example, the use of invasive tracebacks could conceivably bring the world to the brink of nuclear confrontation. Clearly, the moral disvalue in both scenarios outweighs the good to be done by identifying the party ultimately culpable for the attacks.

At this stage, it might be tempting to conclude that these examples show that the use of invasive tracebacks *violates* the Necessity Principle. Since their effects in the worst-case scenarios on intra- and inter-cultural community-building efforts results in significantly more moral disvalue than can be counterbalanced by the moral value achieved by their use, it follows, on this line of reasoning, that the use of invasive tracebacks violates the Necessity Principle.

This reasoning, however, fails to show that, as a general matter, the use of invasive tracebacks violates the Necessity Principle because the use of such technologies in any given instance need not result in the worst-case scenario. Just because an act *can* result in a particular scenario doesn't mean it *will* result in that scenario. Indeed, in any given instance, a party contemplating an active response using invasive tracebacks will have little reliable evidence regarding the probability that the worst-case scenario will result.

Nevertheless, we can justifiably draw the conclusion that, as a general matter, private parties cannot justifiably use invasive tracebacks on the strength of the Necessity Principle – precisely because the probabilities of the worst-case scenario cannot reliably be estimated. Here it is essential to recall what I described as an evidentiary (or epistemic) precondition that must be satisfied in order to justifiably take action under any ethical principle: it is a necessary condition for justifiably acting under an ethical principle that one has adequate reason to believe that its application-conditions are satisfied.

As a general matter, this evidentiary condition will not be satisfied in ordinary situations where private parties are contemplating an active defense involving invasive tracebacks. Insofar as private parties, as a general matter, lack sufficient information to reliably estimate the probabilities of the worst-case scenarios, they lack adequate reason to think that the moral value outweighs the moral disvalue associated with using tracebacks and hence lack adequate reason to think that the application-conditions of the Necessity Principle are satisfied. Thus, absent special knowledge, private parties cannot justify using invasive tracebacks on the strength of the Necessity Principle.

It is true, of course, that the claim that one *cannot justify* using invasive tracebacks by reference to the Necessity Principle is weaker than the claim that the use of invasive tracebacks *violates* the Necessity Principle; but the practical implications are the same. In neither case is it permissible for private parties to use invasive tracebacks under the Necessity Principle. Since, as I have argued, there is no other principle that would justify use of such technologies, it is morally impermissible for private parties to respond to hacker attacks – absent highly unusual circumstances – with invasive tracebacks.

Acknowledgments

This paper was partially supported by funding from Cisco Systems Critical Infrastructure Assurance

Group. The views expressed here are not necessarily those of Cisco Systems. I am grateful to the participants of the First Post-Agora Active Defense Workshop, Seattle, WA (September 12, 2003) for discussion that helped to shape my views on invasive tracebacks. I am especially indebted to Dave Dittrich for patiently explaining the technical issues.

References

Dave Dittrich, Information Assurance Research at the Information School and Senior Security Engineer at Computing and Communications, University of Washington, E-mail, dated, November 29, 2003

Dave Dittrich and Kenneth Einar Himma, "Active Defense," in Hossein Bidgoli (ed.), The Handbook of Information Security, John Wiley & Son, Inc., forthcoming 2005

Kenneth Einar Himma, "Targeting the Innocent: Active Defense and the Moral Immunity of Innocent Persons from Aggression," Journal of Information, Communication, and Ethics in Society*, vol. 2, no. 1 (January 2004)*

Vikas Jayawal, William Yurcik, and David Doss, "Internet Hack Back: Counter Attacks as Self-Defense or Vigilantism?" Proceedings of the IEEE International Symposium on Technology and Scoiety*, Raleigh, NC (June 2002), 5. Available from: http://www.sosresearch.org/publications/ISTAS02hackback.PDF*

Michael Otsuka, "Killing the Innocent in Self-Defense," Philosophy and Public Affairs*, vol. 23, no. 1 (1994), 74-94*

D. Radcliff, "Should you strike back?" ComputerWorld *(November 13, 2000); available from http://www.computerworld.com/governmenttopics/government/legalissues/story/0,10801,53869,00.html*

Proceedings of the symposium "Localizing the Internet. Ethical Issues in Intercultural Perspective" sponsored by Volkswagen*Stiftung*, 4-6 October 2004, Zentrum für Kunst und Medientechnologie (ZKM, Karlsruhe)

---

[i] "Active defense" may be slightly misleading since it suppresses the fact that there is a range of potential responses available to the victim of an attack; Dave Dittrich and I have proposed the adoption of "Active Response Continuum" to call attention to this important feature of active defense. For a comprehensive discussion of the technical, ethical, and legal issues, see Dave Dittrich and Kenneth Einar Himma, "Active Defense," forthcoming in Hossein Bidgoli (ed.), The Handbook of Information Security, John Wiley & Son, Inc., 2005. Nevertheless, I will defer in this essay to existing conventions.

[ii] For example, the host of WTO servers responded to a denial of service (DoS) attack on those servers by redirecting the incoming packets back to the attacking network instead of simply dropping the packets at the router, which would have sufficed to end the attack. See, e.g., D. Radcliff, "Should you strike back?" *ComputerWorld* (November 13, 2000); available from http://www.computerworld.com/governmenttopics/government/legalissues/story/0,10801,53869,00.html.

[iii] State use of active defense raises a very different set of issues, as a morally legitimate state can permissibly do many things that private individuals and entities cannot permissibly do – such as tax and punish private persons and entities.

[iv] I am indebted to Rafael Capurro for making me see the need for this qualification. See *Polylog: A Forum for Intercultural Philosophy* (http://www.polylog.org/index-en.htm) for helpful resources dealing with cultural and intercultural issues in philosophical methodology.

[v] On this imprecise but common usage, a purely theoretical risk is one of such small probability that it can be dismissed from practical deliberations as mathematically insignificant.

[vi] 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.

[vii] Not everyone accepts this view. Michael Otsuka argues that there is no morally significant difference between innocent attackers and innocent bystanders. Both are immunized from infringement of their rights by persons defending against culpable attack by the fact that they bear no moral responsibility for the attack. Otsuka, "Killing the Innocent in Self-Defense," *Philosophy and Public Affairs*, vol. 23, no. 1 (1994), 74-94.

[viii] For an explanation of the distinction between infringing and violating a right, see Note 6 above.

[ix] There is, of course, some vagueness in the notion of "reasonable likelihood." Unfortunately, most ethical principle can be adequately expressed only in language that is vague at the margins. What uncertainty there is about the boundaries of "reasonable likelihood" will not, however, affect the argument I give in this paper.

[x] It is worth noting that the Necessity Principle is a principle of the criminal law of many Western jurisdictions. For example, Section 35.05 of the New York Penal Code provides that "conduct which would otherwise constitute an offense is justifiable and not criminal when … [it] is necessary as an emergency measure to avoid an imminent public or private injury which is about to occur by reason of a situation occasioned or developed through no fault of the actor, and which is of such gravity that, according to ordinary standards of intelligence and morality, the desirability and urgency of avoiding such injury clearly outweigh the desirability of avoiding the injury sought to be prevented by the statute defining the offense in issue." Similarly, section 3.02 of the Model Penal Code provides that "[c]onduct that the actor believes to be necessary to avoid harm or evil to himself or to another is justifiable, provided that … the harm or evil sought to be avoided by such conduct is greater than that sought to be prevented by the law defining the offense charged."

[xi] For Hill's tragically misguided views, see http://www.armyofgod.com/PHillonepage.html.

[xii] Indeed, this is a very consequence of the claim that a fetus is a moral person. If a fetus has a full and equal set of moral rights, then murdering a fetus violates the same right of life that murdering a newborn infant violates and is just as grave a moral offense. This is why the issue of fetal personhood is so crucial to the abortion debate.

[xiii] Indeed, it is for these reasons that criminal cases are prosecuted by the state instead of the individual. In civil cases, it is entirely up to the victim to decide whether she wishes to seek compensation and to initiate the legal steps that would result in an appropriate court order; since only the individual victim of a civil wrong has a compelling claim for compensation, the individual victim has discretion to prosecute her own lawsuits as plaintiff. In criminal cases, it is the state that decides whether to pursue criminal charges against an offender.

[xiv] Notably, the same is true of a situation in which the innocent party is home at the time and can be asked by the security officer for her permission to come onto the land. It seems clear that the security officer would be justified in coming onto the land even if the innocent party refused her permission. While the infringement of the innocent party's property rights in this case is specifically intended, the infringement is so small relative to the great moral good it accomplishes that it does not constitute a violation; the innocent party's legitimate interests in her property do not include authority to deny its use in such circumstances.

[xv] Vikas Jayawal, William Yurcik, and David Doss, "Internet Hack Back: Counter Attacks as Self-Defense or Vigilantism?" *Proceedings of the IEEE International Symposium on Technology and Scoiety*, Raleigh, NC (June 2002), 5. Available from: http://www.sosresearch.org/publications/ISTAS02hackback.PDF.

[xvi] *Id.*

[xvii] Email from Dave Dittrich, Information Assurance Research at the Information School and Senior Security Engineer at Computing and Communications, University of Washington, November 29, 2003.

[xviii] I am assuming that nations are fairly characterized as "cultures." These, of course, are not the only cultures; there are a variety of cultures that are located inside national boundaries and that transcend them. I am grateful to Rafael Capurro for pointing this out to me.

[xix] D. Radcliff, "Should you strike back?" *ComputerWorld* (November 13, 2000); available from http://www.computerworld.com/governmenttopics/

government/legalissues/story/0,10801,53869,00.html.