Active Response to Computer Intrusions

David Dittrich and Kenneth Einar Himma, University of Washington

INTRODUCTION: THE CONCEPT OF ACTIVE RESPONSE

The active response continuum defines a category of digital response to unauthorized digital intrusions1 and hence falls within a wide spectrum of potential responses by private entities.2 At one end of the spectrum is the wholly passive, unknowing victim who relies entirely on just the inherent capabilities of the software that comes with the computer he or she purchased and who does not know when it is being attacked. At the other end is the active, fully engaged victim who deliberately pursues a series of discrete tactics with a set of well-defined objectives in mind. As the term "active" indicates, active response measures fall toward the latter end of the spectrum.

As defined in this chapter, measures falling within the active response continuum have the following characteristics. First, these measures are, of course, digitally based; physically assaulting someone who is committing a digital trespass would not be an instance of active response. Second, they are reactive in the sense that they are implemented following detection of an unwanted digital trespass (e.g., scanning ports) that are not fairly characterized as hacking. Third, they are noncooperative in at least the minimal sense of being implemented without the consent of at least one of the parties involved in or affected by the intrusion. Finally, they usually have causal impacts on remote systems (i.e., those owned or controlled by some other person). These tactics range from more benign information-gathering measures (e.g., tracebacks) that have an impact on remote systems without impairing their ongoing operations and functions to more aggressive measures (e.g., denial of service counterattacks) expressly intended to inhibit or even stop the operations and functions of remote systems.

At an intuitive level, the definition of "active response" attempts to pick out digital acts that would be characterized as hacking if performed without provocation. Theorists and policymakers have become concerned with the phenomena discussed in this chapter precisely because of their resemblance to hacker attacks; it is, for this reason, that active response measures are sometimes referred to as "counterhacking" or "hacking back." Although the definition of "active response" above picks out some acts (e.g., scanning ports) that are not fairly characterized as hacking, even those acts are minimally intrusive. Indeed, although they might very well be morally justified all things considered, they raise, at least initially, the same sorts of privacy and property concerns that are raised by acts that are fairly characterized as hacking.

The definition above also attempts to incorporate the idea that these measures are taken in response to an unauthorized intrusion. The idea that such measures are reactive implies that measures intended to detect the occurrence of an intrusion do not fall within the active response continuum. Reactive measures, as defined above, are deliberately contrived as a response to an intrusion that has previously been detected—though it is true, of course, that detection and response might sometimes proceed together as the intrusion continues. Although many measures taken in response to computer intrusions fall within the active response

---

1 Other terms used to pick out such measures are "active defense," "counterhacking," "hacking back," etc.
2 This chapter is limited to private intrusion responses. The response by public entities to digital intrusions raises very different technical, ethical, and legal issues.
continuum, not all do. Measures reasonably calculated to stop either an ongoing attack or the harm it is causing are typically characterized as defensive; however, active response measures can serve a variety of purposes that are not, strictly speaking, defensive in either of these respects. For example, information-gathering efforts can be related to efforts to adopt measures calculated to stop an attack or its harmful effects, but they can also be directed at providing law enforcement agencies with sufficient evidence to prosecute culpable parties. In addition, some measures sometimes adopted for defensive purposes, they regard for whether such measures actually bring about its cessation. Accordingly, although active response tactics are sometimes adopted for defensive purposes, they are also frequently adopted for investigatory and offensive purposes— and, indeed, can be employed in ongoing information warfare.

Thus conceived, measures falling within the active response continuum are compatible with the efforts of law enforcement agencies to investigate and prosecute computer crimes. As is readily apparent, efforts by private victims to gather and preserve information about the attack can assist law enforcement efforts to investigate and prosecute an attack. Such information can provide not only helpful investigatory leads but can also form the foundation for the evidentiary base needed to prosecute culpable parties successfully.

Even so, it is worth emphasizing at the outset that active response measures are increasingly adopted by private firms as a substitute for involving law enforcement agencies. There are a variety of reasons for this practice. First, the resources available to law enforcement agencies for responding to digital intrusions have simply not kept pace with the frequency and severity of digital intrusions. The perceived increasing rate of law enforcement has led to a sense among private victims that it is far more efficient to respond without involving law enforcement. Second, and equally important, many commercial victims worry about the effects that publicizing an attack might have on their relationships with customers. The concern is that their customers would become alarmed after learning of security breaches and would ultimately respond by taking their business elsewhere. Therefore, such firms believe that the best way to minimize the risk of publicity and such deleterious effects is to respond internally to digital intrusions without involving law enforcement.

This chapter provides an overview of the active response continuum and considers the morality and legality of such tactics. It defines various levels of intrusion response in relation to the level of the victim's conscious involvement and posture, describes various technical barriers to active responses and the various ways in which law enforcement might be involved, characterizes active response measures along a spectrum ranging from benign to aggressive measures, and considers the ethicality and legality of responses falling along that spectrum.

It is important to keep in mind that the topic of active response is a novel one, with an academic and popular literature that is exceedingly small compared to the literature available for other security-related topics. As such, many of the descriptive and normative issues are currently being worked out by theoreticians and practitioners. As one might expect, then, theoreticians and practitioners disagree on a number of these issues. Such disagreements are not limited, however, to the usual ethical disagreements; they also involve significant problems of classification that seek to distinguish the various levels and characteristics of active response.

For this reason, the chapter in this chapter should not be considered the final word on any of the issues considered below. As the topic of active response continues to attract more interest from theoreticians, ethicists, and ultimately lawmakers, one should expect that new descriptive taxonomies will emerge that may well change the normative landscape— changes that undoubtedly will result in changes in the content of existing law. Even so, this chapter is intended to provide a plausible conceptual and normative foundation for understanding the descriptive, ethical, and legal issues of active response and to enable the reader to follow the conversation as it evolves.

LEVELS OF INTRUSION RESPONSE

As a first step toward understanding active response, it is important to get a sense of the range of potential responses to digital intrusions. At least five different levels of intrusion response can be distinguished along a spectrum according to the degree of the victim's deliberative engagement. The spectrum begins at Level 0 where the victim has almost no knowledge of the intrusion and is hence wholly unengaged and ends at Level 4 where the victim is fully engaged and is acting independently of other involved parties (Table 1). Active response occurs primarily at this fourth level, which is itself divided into sublevels defining a subcontinuum that progresses from less intrusive to more intrusive, less risky to more risky, and less disruptive to more disruptive sublevels of involvement.

Level 0—Unaware

The defining attribute of a Level 0 response is the utter lack of involvement of the attack victim. The victim has no knowledge of the intrusion and takes no action whatsoever. At Level 0, the victim (construed to include both the owner/operator of the computer system and the organization to which he or she reports) takes no active role in responding either directly or indirectly to an ongoing attack. Indeed, at Level 0, the victim does not do so much as

3 For a general discussion of information warfare, see Wireless Information Warfare, in this handbook.

5 We are indebted to attendees at the First Agora Workshop on Active Defense for this taxonomy and supporting analyses. The Agora, an information security group in Seattle, Washington, founded and led by Kirk Bailey (currently CISO for the City of Seattle), has held three workshops on the topic of active response, the first of which was held in June 4, 2001. Much of the terminology and description of the active response continuum here was derived by workshop participants through the assistance of Captain Jake Schaffner (USN Ret.), who co-moderated this first workshop with Nick Mullis.
victim does not actively respond in any thoughtful way to a virus is detected. They are passive in the sense that the victim receives little input from the operations of these resources.

This category of user is a common target for attackers wishing either to (1) retain anonymity through the use of stepping stones or proxies; (2) install malicious software that permits remote control of the computer for use in distributed attacks or intrusions (e.g., distributed denial of service attacks or distributed spam transmission); or (3) steal disk space for storage of stolen content (e.g., pirated software, data stolen from other compromised hosts, and cache malicious software).

Level 1—Involved

Level 1 responses involve minimal engagement on the part of the victim. The victim establishes (either directly or via proxy) a day-to-day defensive posture involving only resources within his or her ownership or operational control. Such resources may include (1) use and maintenance of personal firewall products; (2) use of personal firewall software to limit remote access to services installed as part of the operating system (e.g., Windows System Message Block and NetBIOS name services); or (3) use of a hardware firewall between the network attachment point and hosts in a local area network (e.g., a home LAN on a broadband cable). At this level, there is little interaction between the victim and the operations of these resources.

Level 1 responses are fairly characterized as passive, prophylactic, and silent. They are prophylactic in the sense that they are intended primarily to prevent attacks and do little to respond in an efficacious manner when an attack is detected. They are silent in the sense that the victim receives little input from the operation of these resources; for example, a victim at this level might not even have set up an antivirus program to alert him or her when a virus is detected. They are passive in the sense that the victim does not actively respond in any thoughtful way to the discovery of an intrusion. Even in instances where antivirus software alerts the victim to a threat of some kind, his or her response at this level is usually limited to accepting, without any significant deliberation, the recommendations made by the anti-virus software for cleanup.

Level 2—Interactive

This level is characterized by the beginning of an active engagement with the threat. At Level 2, the victim responds to evidence of an intrusion by taking minimally deliberative measures to modify resources under his or her ownership or control. A minimal Level 2 response, for example, may involve adjusting the security settings on a personal firewall to respond to a specific kind of detected threat—perhaps configuring software to ignore pings. A more fully engaged Level 2 response may involve looking up domain registry information for the Internet protocol (IP) of the "attacking system" and reporting the attempted intrusion or actual compromise to the site that owns this IP. In some cases, the victim at this level may go so far as to report these events to incident coordination sites, such as CERT/CC, and/or to law enforcement.

Though Level-2 responses actively engage the threat in some way, the extent of engagement is comparatively unsophisticated. These victims make little effort to investigate the intrusion or to perform a forensic analysis of compromised hosts, and are content to rely on others to take action (e.g., the victim's incident response team or law-enforcement agencies). At this level, victims typically respond to a successful intrusion by formatting the drive and re-installing the operating system.

Level 3—Cooperative

Level 3 is distinguished from lower levels in that it involves an attempt on the part of the victim to reach out beyond those resources owned or operated by him or her. At Level 3, the victim attempts to enlist the cooperation of other organizations/systems in taking joint measures intended to attribute, mitigate, or eliminate the threat. It is important to note that the causal effects of action at

---

Table 1 Levels of Intrusion Response

<table>
<thead>
<tr>
<th>Level</th>
<th>Victim Posture</th>
<th>Characteristic Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Unaware</td>
<td>None: Passive reliance on inherent software capabilities</td>
</tr>
<tr>
<td>1</td>
<td>Involved</td>
<td>Uses and maintains antivirus software and personal firewalls</td>
</tr>
<tr>
<td>2</td>
<td>Interactive</td>
<td>Modifies software and hardware in response to detected threats</td>
</tr>
<tr>
<td>3</td>
<td>Cooperative</td>
<td>Implements joint traceback, cease-and-desist measures,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(active response) and retaliatory countermeasures</td>
</tr>
<tr>
<td>4</td>
<td>Noncooperative</td>
<td>Implements invasive traceback, cease-and-desist measures,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(active response) and retaliatory countermeasures</td>
</tr>
</tbody>
</table>

---

4 Noncooperative Implements invasive tracebacks, cease-and-desist measures, or technologies. These na"ıve users pay little, if any, attention to patches, antivirus software, settings on Web browsers, and have no comprehension of plaintext password and other vulnerabilities. Such users lack the ability to discover when their computers are compromised by worms or viruses. In the event of an intrusion of some sort, they simply continue to operate their computers as usual.

An example of a Level 0 response is the typical home-broadband customer who purchases a computer through a retail outlet and knows very little about security issues or technologies. These naive users pay little, if any, attention to patches, antivirus software, settings on Web browsers, and have no comprehension of plaintext password and other vulnerabilities. Such users lack the ability to discover when their computers are compromised by worms or viruses. In the event of an intrusion of some sort, they simply continue to operate their computers as usual.

5 Typically, the attacking site is the last hop in a potential chain of stepping stones located at a distance from the culpable source of the attack. 7 On its Web page, CERT/CC is described as a "major reporting center for Internet security problems. Staff members provide technical advice and coordinate responses to security compromises, identify trends in intruder activity, work with other security experts to identify solutions to security problems, and disseminate information to the broad community" (see http://www.cert.org/).
Noncooperative intelligence gathering can cause conflicts among victim sites. Suppose, for example, that site A takes it upon itself to scan the network of site B, attempting, say, to identify back doors or similar systems to those compromised at site A. Even if A attempts to cooperate with B by providing it with the results of the analysis to help B improve its understanding of the attack, B might still consider A’s action objectionably intrusive if B feels its privacy has been threatened.

Such measures can also have the unintended effect of providing useful information to the attacker. If the attacker notices that A is scanning an attacked system, the attacker may respond by attempting to eliminate evidence of his or her presence in the system by deleting files and re-initializing hosts being used as stepping stones (including routers.) The effects of noncooperative measures in such circumstances not only make it more difficult to identify the attacker but can also result in further damage to the compromised system if the attacker deletes valuable files.

Noncooperative Cease-and-desist Measures
Noncooperative cease-and-desist measures are contrived to stop an attack by rendering the attacking machines inoperable and hence, unlike the measures previously discussed, are intended to have a causal impact on someone else’s computers in a direct and functional way. For example, a victim might attempt to shut off distributed denial of service (DDoS) agents by using vulnerabilities in those programs to inject commands or by triggering denial of service (DDoS) agents by using vulnerabilities in those programs to inject commands or by targeting known vulnerabilities (e.g., the Windows RPC/DCOM vulnerability exploited by the MSBlaster and Nachi worms) in ways that cause infected systems to crash. Noncooperative cease-and-desist measures, then, are expressly contrived to impair the operation and functioning of remote systems, causing them to many more intrusive way than the measures discussed previously.

Retribution or Counterstrike
Retributive or counterstrike measures are fairly characterized as the most aggressive active response tactics. As the name suggests, these measures are contrived to retaliate against attacking machines by inflicting something that is likely to be perceived as damage or a harm of some kind by the attacker (or unknowing agents). The structure of a counterstrike may mirror the structure of the attack or it may differ in key respects. However, in every case the immediate point is to inflict harm and is usually motivated ultimately by a desire to either stop or punish the attack. Like noncooperative cease-and-desist measures, retributive counterstrikes are intended to have a causal
Public preemptive response poses a host of practical difficulties. The state must have sufficient intelligence capabilities to be able to detect and identify potential attacks accurately before they are staged. These capabilities are, of course, very expensive to maintain and deploy, but also have a variety of other social costs. For example, the public may perceive a decrease in its personal privacy without a reasonable increase in security against terrorist actions. Further, law enforcement agencies may begin arresting the wrong people, negatively affecting the overall law enforcement effort, as well as diminishing the public’s sense that it is treated fairly and justly by law enforcement agencies.

POTENTIAL TECHNICAL BARRIERS FOR INTRUSION RESPONSE

Several factors can complicate efforts to implement an effective response to a digital intrusion. First, digital evidence is somewhat less durable than material evidence, complicating efforts to understand an attack and identify its ultimate source. Second, the victim must understand the structure of an attack to determine an appropriate response. Third, digital attackers have a variety of sophisticated tools for effectively concealing their identities and frustrating efforts by victims to understand an attack.

Volatility of Digital Information

In Guidelines for Evidence Collection and Archiving, Dominique Brezinski and Tom Killalea (2002) elaborate several “best practices” for collecting information during an intrusion that take into account the durability of digital evidence. In particular, they recommend attempting to collect information according to “order of volatility”, that is, collecting information in order of durability starting, of course, with the information that is least durable. Brezinski and Killalea characterize the following sources of information in a hierarchy from most to least volatile:

- registers, cache
- routing table, ARP cache (host route information), process table, kernel statistics, memory
- temporary file systems
- disk
- remote logging and monitoring data that are relevant to the system in question
- physical configuration, network topology
- archival media.

Understanding this order is important in considering an active response because much of the data needed by the victim to understand an attack are extremely volatile. For example, an analysis of network flows and actual packets, which are at the top of the hierarchy, is crucial to understanding the structure and methodology of an intrusion.

---

9 Prudential considerations are distinct from ethical or legal considerations—though they may overlap. What is in one’s self-interest may not be either ethical or legal, and conversely.

10 A Scotland Yard detective has stated that “small groups of young people creating a resource out of a 10,000- to 30,000-strong computer (bot) network are renting them out to anybody who has the money” Scotland Yard, (2004).

11 These cost no other comparable laws at the time of setting, and it is not clear under what circumstances a justification could be made that taking preemptive action against an attacker by government or private sector owner/operators was warranted.
Understanding Attack Methodology

A victim of a digital attack cannot mount an appropriate response of any kind beyond Level 1 without having some understanding of the attack’s structure and methodology. Even an effective interactive response (Level 2) requires some understanding of the basic characteristics of the attack; a victim cannot modify software and hardware in a genuinely responsive way without understanding the attack. For example, a defender may have two options for obtaining a piece of information about the attack, one that is detectable by the attacker and one that is not. If the defender is not aware that there are two options or unknowingly chooses the detectable option, thinking it to be the only appropriate one, the attacker may learn of the defender’s pursuit and avoid the defensive action, return through remaining back doors, and possibly do significant damage as he or she “bugs out” of the network.

Attribution

Preparing an active response that is both appropriate and likely to be efficacious depends on identifying both the immediate (i.e., innocent agents) and ultimate (i.e., culpable attacker) sources of an attack; that is, it depends on an accurate “attribution” of the attack to all the parties that are responsible either immediately or ultimately for the attack. Accurate attribution is obviously important because the responses by the military, the intelligence community, federal law enforcement, and the business community will differ radically depending on who is responsible for the attack.

Unfortunately, attribution can be extremely difficult. First, there are always a large number of possible culprits in any substantial Internet attack. The ultimate source of an attack could be (1) an unskilled “script-kiddie” using ready-made tools he or she downloaded from the Web, (2) a “black hat by night” employee of a large security company trying to increase market demand for his or her company’s products, (3) a nation-state actor developing and testing an information warfare capability, (4) a spammer’s hired gun who is trying to attack those who want to end spam or who is trying to establish a massive spam delivery network of compromised hosts, or (5) a terrorist organization trying to find a way to affect the economy by disrupting important businesses. Second, reasonably sophisticated attackers typically conceal their identities by using stepping stones, open wireless access points, stolen credit card numbers, cell phones with built in modems, and dial-in services without caller-ID. These factors make reliable traceback and accurate attribution extremely challenging.

LEVELS OF FORCE: BENIGN THROUGH AGGRESSIVE RESPONSES

This section classifies various intrusion responses according to levels of “force.” Although there are clear differences between physically forceful actions and digitally intrusive actions, there are sufficient similarities to justify characterizing the latter as involving force of some kind. Despite the probable limits to the analogy between digital and physical force, these limits are of little importance here and can be disregarded safely. Table 2 shows the levels of force.
Benign

Benign activities are those involving operations that have no direct causal effects on remote systems and are not adopted from a self-consciously noncooperative posture or attitude. Such measures include operations intended to gather information, as well as to address or correct vulnerabilities in the victim’s networks. Most benign measures do not fall within the active response continuum as that concept was defined above, but some do—for an interesting reason. Some benign measures potentially affect the legitimate interests of other persons even if, strictly speaking, they do not have a causal impact on remote systems. Third parties may have legitimate privacy interests implicated by measures that have a causal impact on only those systems owned or controlled by the victim. If, for example, the victim is an Internet service provider renting Internet access and storage to a third party, measures that have a causal impact on only those resources within the victim’s ownership or control may affect the legitimate moral interests of the victim. Insofar as the victim takes unilateral action without obtaining consent from interested third parties in situations in which consent is required ethically or legally, such action is fairly characterized as “noncooperative” as opposed to “uncooperative.” Although the latter term presupposes a deliberate refusal to behave in a cooperative way, the former notion does not. Thus, behaviors that are not cooperative in circumstances in which cooperation is required are “noncooperative” even if the failure to cooperate is unknowing. Unilateral benign measures that potentially infringe the legitimate interests of third parties, then, are fairly characterized as noncooperative and hence as active response.

Sniffing

Sniffing (i.e., monitoring of network traffic) can occur only on LAN segments. This requires access to a device on the LAN on which you wish to sniff (e.g., connected to a wall port in a switched network, or ability to associate with a wireless access point). Attackers using sniffers typically take over a computer and use its interface to promiscuously capture all traffic that is accessible to that host. Using techniques that manipulate link-level traffic management functions, such as ARP cache poisoning and MAC table overflowing in switches, switches and virtual local area networks (VLANs) can sometimes be bypassed to increase the traffic that can be sniffed. Routers themselves are sometimes compromised and used to sniff traffic as well. Laws that prohibit monitoring of electronic communications typically have exemptions for owners and operators of networks or telecommunications systems and for those with authority for the investigation of breaches in network security; these exceptions allow such persons to monitor traffic. The main requirement is that the activity be done for protection of the network and computers involved.

Readdressing Hosts or Networks

Attackers commonly attempt to map out network infrastructure to determine how to achieve their objectives inside the target network—a process that can take months or even years to complete if an attacker is trying to do it with stealth (a so-called low and slow scan.) If the victim can rapidly readdress hosts and devices on its network after the commencement of an attack, the victim can effectively blind the attacker and force him or her to rescan the network. This may enable victims to detect the systems from which attackers are implementing scans and to identify hosts that are being used as stepping stones.

Deception Using Honeypots

Honeypots, as commonly defined, are computer resources installed for the purpose of being compromised (see The Use of Deception Techniques: Honeypots and De- coys in this Handbook; http://project.honeynet.org/papers/honeynet/index.html). Honeypots can be used to augment existing firewalls and intrusion detection systems to identify malicious activity. Some honeypots not only appear to provide services desired by an attacker, but can also mimic the operating system fingerprint that results when an attacker scans the network with certain tools. In that attack situation, a defender can readdress certain hosts on the network, configure a system running those honeypots to replace the existing systems, and make it look like there are many more on the network just like it.

Table 2

<table>
<thead>
<tr>
<th>Level</th>
<th>Causal Impacts</th>
<th>Characteristic Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign</td>
<td>Limited to victim’s own systems</td>
<td>Sniffing, scanning, readdressing hosts, honeypots</td>
</tr>
<tr>
<td>Intermediate</td>
<td>Affects remote systems but not calculated to produce damage</td>
<td>Invasive tracebacks, remote evidence collection</td>
</tr>
<tr>
<td>Aggressive</td>
<td>Impacts calculated to produce damage in remote systems</td>
<td>Remote exploitation, corruption of data, denial of service</td>
</tr>
</tbody>
</table>

Scanning

The attacker uses scanning to learn about the topology of the network and the devices being used, and the victim can do the same. Understanding the vulnerabilities being exploited by an attacker is essential to an effective response; scanning can help the victim find vulnerable and compromised hosts, thereby facilitating cleanup and evidence preservation. Indeed, a victim can continually scan his or her network to build an historical database of operating systems, listening ports, and services that can be queried on demand when investigating an attack to find all hosts matching a known exploitable profile. By performing a continual comparison of this same information, changes over time can be detected, which may signal intrusion activity. There are passive ways of gaining this same

---

**Table 2** Levels of Force

<table>
<thead>
<tr>
<th>Level</th>
<th>Causal Impacts</th>
<th>Characteristic Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign</td>
<td>Limited to victim’s own systems</td>
<td>Sniffing, scanning, readdressing hosts, honeypots</td>
</tr>
<tr>
<td>Intermediate</td>
<td>Affects remote systems but not calculated to produce damage</td>
<td>Invasive tracebacks, remote evidence collection</td>
</tr>
<tr>
<td>Aggressive</td>
<td>Impacts calculated to produce damage in remote systems</td>
<td>Remote exploitation, corruption of data, denial of service</td>
</tr>
</tbody>
</table>
intelligence through monitoring network flows, although this monitoring only identifies hosts that are actively “talking” on the network.

**Session Hijacking/TCP Session “Sniping”**

There are tools that enable a victim to hijack and terminate established transmission control protocol (TCP) sessions as well as from being established. The careless and sporadic use of such tools is, however, likely to be noticed by the attacker—though an attacker could always misinterpret the results as having been caused by intermittent network failures. However, it is important to note that these tools are generally efficacious only if used on the same LAN segment as the hosts the attacker is using.

**Intermediate**

Activities in this category involve causal interaction with remote systems outside a defender’s network, but are neither intended nor reasonably likely to cause harm to those systems. It is worth noting that noncooperative measures falling in this category are fairly characterized as active response as this notion was defined above—though, again, the reader is cautioned not to draw any substantive normative conclusions merely on the strength of a characterization of something as involving, or not involving, active response. The above taxonomy can assist a normative analysis, but is no substitute for such an analysis.

**Following Attack Paths in Reverse**

Victims can attempt to follow attack paths back to the ultimate source of an attack. If a victim knows (1) the attacker’s methodology for using back doors and exposed network services for establishing stepping stones or proxy relays, (2) the passwords and/or account names favored by the attacker, and (3) the IP addresses used for entry/exit from the victim’s network, he or she can follow the attacker’s trails backwards through the network. Tracing attack paths through systems within the victim’s own network can be done without much worry if there are policies established. The careless and sporadic use of such tools is, however, likely to be noticed by the attacker—though an attacker would simply be restarted as a normal course of actions if, among other things, relevant to an ongoing criminal investigation. Aggressive actions defined here as aggressive are reasonably likely to interfere with the availability, integrity, confidentiality, or authenticity of information systems outside one’s own network. Aggressive measures include those intended or highly likely to result in something that the target would regard as harm or damage. Aggressive actions are all fairly characterized as active response since, as a conceptual matter, only a noncooperative act can be aggressive.

**Remote Exploitation**

One class of aggressive measures involves victim penetration and exploitation of remote systems. Understanding a remote system’s vulnerabilities allows the victim to penetrate those systems using exploit programs (i.e., programs that take advantage of vulnerabilities in a computer system to gain access.) These programs may include exploits used by the attacker within the victim’s own system.35

There are several serious risks here that counsel against use of such measures. First, remote exploitation can cause major disruption to the host if the exploit crashes servers, network stacks, or the operating system itself. Second, it can leave digital tracks that the attacker would notice. Third, it can be detected by the site’s incident handlers and treated as though it were a separate computer crime.

**Corruption of Data**

Another class of aggressive measures involves alteration of data being used by the attacker. In a case where it is known that the attacker is using a file system cache to store sniffer or vulnerability scanner logs, an effective response can involve the targeted editing or destruction of some or all of these files. Because it can take significant time for an attacker to successfully identify new hosts to attack, such measures can force the attacker to return to certain hosts or networks. At the very least, they can delay the attacker for long enough to allow the victim to finish an analysis of malware artifacts or system images and to gather new network flow information. However, it is important to realize that these actions can also destroy evidence that could be used in a criminal prosecution and may themselves violate criminal prohibitions on digital intrusions.

**Disabling Services on Remote Systems**

Knowledge of services being used by the attacker on remote systems, together with the possession of accounts/passwords or knowledge of remotely exploitable vulnerabilities in these services, can enable a victim to disable them. The risks here are that (1) disabling the service could have a negative impact on the host or its users, (2) the service would simply be restarted as a normal course of system operation (e.g., through a monitoring daemon, or verified their integrity from known trusted public sources. The USA Patriot Act, however, allows law enforcement to monitor and use such communications if, among other things, relevant to an ongoing criminal investigation.36

Victims should always have a healthy distrust of programs in an attacker’s possession, unless they have thoroughly reverse engineered them or verified their integrity from known trusted public sources.

---

35 In the U.S., for example, the Electronic Communications Privacy Act prohibits “interception and disclosure” of certain “wire, oral, or electronic communications.” See 18 U.S.C., Section 2501 et seq. The USA Patriot Act, however, allows law enforcement to monitor and use such communications if, among other things, relevant to an ongoing criminal investigation.

36 Victims should always have a healthy distrust of programs in an attacker’s possession, unless they have thoroughly reverse engineered them or verified their integrity from known trusted public sources.
Denial of Service
another possibility is to use remotely exploitable vulnera-
abilities to cause hosts or the network infrastructure itself
to fail at the other site, thereby taking control of remote
systems out of the hands of the attacker, or overwhelming
the network bandwidth of the site to the same end. The
effect of such measures is to deny users access to the con-
tents of the other site and hence to “deny service” to them.
Denial of service (DoS) responses are highly aggres-
sive measures that are problematic for several reasons
(see Denial of Service in this Handbook). First, such re-
sponses can have unforeseeable catastrophic effects on
remote sites that are not unreasonably characterized as
“collateral damage.” For example, it might be likely that
a DoS response could cause physical damage to a system
involved in critical patient care, potentially resulting in
risk of harm or even in the death of a patient.18 Similarly,
it could be that a DoS response might result in financial
losses that well exceed the value of the resources the victim
is attempting to defend and in the worst-case scenario re-
sult in the losses of jobs of innocent persons. Second, it is
rarely clear that denial of service will accomplish the goal
of removing access to the compromised systems. In large-
scale attacks where the attacker controls tens or hundreds
of servers, the attacker typically lacks the abil-
ity to deny service to all of them.18 Third, there is the pos-
sibility that the counterattack would simply be filtered out
somewhere between the attacking network and the target
networks. Not only will the counterattack fail as a means
of removing control of these systems by the attacker, it may
even disrupt the victim’s own network connectivity as a result.

THE ETHICS OF ACTIVE RESPONSE
This section is concerned with whether it is ethically
permissible for private parties to adopt the various ac-
tive response measures described above in response to
a digital attack. The structure of the analysis is as
follows. The first section attempts to identify each of
the substantive ethical principles potentially relevant in
evaluating whether a particular active response strategy
is permissible. The second section identifies an additional
ethical principle that states an evidentiary precondition
for justifiably acting on other substantive ethical princi-
pies (such as those identified in the first section). Finally,
the third section applies these principles to aggressive, in-
termediate, and benign active response measures.19
One preliminary observation is in order here. As is
standard in the area of applied ethics, the analysis here
does not presuppose any particular general ethical the-
ory like utilitarianism or Kantianism. Instead, the anal-
ysis purports to be grounded in general principles, and
specific case judgments figure prominently in ordinary
ethical judgments and practices. Accordingly, the analy-
sis paper begins by identifying ethical principles that are
commonly accepted in Western industrialized nations20
and proceeds by attempting to identify the implications
of those widespread commitments with respect to the var-
umous levels of active response.

Relevant Ethical Principles
Allowing Force in Defense of Self and Others
It is generally accepted in Western nations that a person
has a moral right to use proportional force when neces-
sary to defend against an attack. If, for example, A is shoot-
ing at B without provocation and B cannot save his or her
own life without shooting A, it is permissible, according
to ordinary judgments, for B to shoot A. If, however, A starts
hitting B without provocation, it would be impermissible
for B to shoot at A; because B’s right of self-defense is lim-
ited to directing proportional force at A, it is permissible for
B only to hit A.

The first ethical principle considered here, then, is a
familiar one that allows a person to use proportional force
when necessary to defend against an attack:

The Defense Principle: It is morally permissi-
ble for one person to use force to defend him- or
herself or other innocent persons against an at-
tack 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 is reason-
ably likely to harm only those persons who are
responsible for the attack.

Although the term “force” has traditionally been used
to describe violent physical attacks in which one person

---

19 There are two reasons for structuring the analysis this way. First, each of
the various levels of active response must be evaluated under each of
the ethical principles. Second, each of the various substantive principles is
qualified by the evidentiary precondition discussed in the second section.
For this reason, the analysis must first identify all of the relevant
substantive and evidentiary principles before considering how any one
might apply to some active response strategy.

20 Here it is worth noting that the principles identified here are incorpo-
rated into the law of every Western industrialized nation. To the extent that
most people accept those laws as legitimate, it is reasonable to conclude
that most people believe they are just—and hence reflect the content of
morality.
attempts to inflict physical harm on another person, it is reasonably construed here as applying to both physical and digital attacks.

Each of the elements of the Defense Principle states a necessary condition for the use of force. First, it justifies only force that is proportional to that used in the attack. Second, force must be necessary in the sense that the victim cannot stop the attack or prevent further harm to him or herself without resorting to force. Third, the Defense Principle justifies the use of force only against persons directly responsible for the attack. Although some (but by no means all) theorists believe this element allows the use of force against an attacker who is innocent of wrongdoing (perhaps because he or she is clinically insane), all agree that the Defense Principle does not justify force against an innocent bystander.

At this juncture, it is worth noting that the Defense Principle will hence justify, at most, forceful active response measures directed at the owners of innocent agent machines; such machines are fairly characterized as "innocent attackers" rather than "innocent bystanders." Active response measures that have significant impacts on innocent bystanders will likely face difficulties under the Defense Principle.

Allowing Otherwise Wrongful Acts to Secure the Greater Moral Good

It is also generally accepted in Western nations that moral- ity allows the enforcement (as opposed to violation) of an innocent person's rights when necessary to secure a significantly greater good.21 For example, if A must enter onto an innocent person's property when necessary to secure a significant moral value, it is morally permissible for A to do so. Though such an act constitutes a prima facie trespass and hence infringes B's property rights, it does not violate B's property rights because it is morally justified.

There are four considerations that explain this judgment. First, stopping a dangerous murderer from escaping, it is morally permissible for A to do so. Though such an act constitutes a prima facie trespass and hence infringes B's property rights, it does not violate B's property rights because it is morally justified.

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.

Like the Defense Principle, the Necessity Principle is construed here as applying in the context of both physical and digital attacks.

The Necessity Principle augments the Defense Principle by allowing acts that 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. However, insofar as the Necessity Principle requires the achievement of a significantly greater good, it does not allow a person to direct at an innocent bystander force that is fully proportional to the force of the attack.

The Necessity Principle is, thus, dissimilar to the Defense Principle in one respect that is significant for the evaluation of active response. Unlike the Defense Principle, the Necessity Principle potentially allows active response measures that have significant impacts on innocent bystanders. Even so, it is clear that the Necessity Principle will allow impacts on innocent bystanders only if the moral value of using the relevant active response measure significantly outweighs the moral disvalue of such impacts.

Punitive or Retaliatory Principles

It might be thought that victims of an attack have a moral right to retaliate against or punish their attackers by inflicting a morally 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 ethically permissible, in this view, for B to catch A and then hit him back in the face. B's retaliatory act is justified because it gives A what A deserves and thereby restores the balance of justice that was disturbed by A's morally wrongful act. Applied to the present context, such an analysis would permit the victim of a digital attack to respond with force as a means of "evening the score." Nevertheless, it is generally accepted that, in any society with a morally legitimate government, it is ethically impermissible for citizens to punish or retaliate against wrongdoing. Mainstream political theorists are unanimous in holding that it is the province of government—and not the individual—in such societies to punish wrongdoers after they have been found guilty in a fair trial. Indeed, vigilantism is universally condemned as wrong; so long as the state is reasonably effective in prosecuting and punishing wrongdoing, citizens are morally prohibited from forceful self-help.22 As a general matter, it is

---

21 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 makes no moral judgment as to whether 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.

22 Such an analysis is presented by Jayawardena, Vrisc, and Moss (2002). The assumption is that self-defense is legitimate, but not retaliatory measures (which are implicitly condemned as "vigilantism").
wrong for private victims to even the score by retaliating or punishing attackers.\textsuperscript{52}

\textbf{An Evidentiary Restriction for Justifiably Acting Under Ethical Principles}

Most theorists and laypersons agree that we have a duty to ensure we have correctly identified the facts and applications of a relevant principle before taking action against a person. Suppose, for example, that A believes without good reason that B has wronged A in some way. If A takes action against B under a principle $P$ without having a minimally adequate reason for thinking that $P$ applies (say, because A lacks a minimally adequate reason to think that B has committed a wrong), then A has committed a wrong against B. A has a duty to be at least minimally justified in believing that $P$ governs the situation; if A does not satisfy this duty, then A must give B the benefit of the doubt before acting on $P$ until A has better evidence that $P$ does, in fact, apply.

There is thus a third principle that is relevant with respect to evaluating active response measures—one that is evidentiary or, as theorists of knowledge put it, “epis- temic” in character:

\textbf{The Evidentiary Principle:} It is morally permissible for one person $A$ to take action under an ethical principle $P$ only if $A$ has adequate reason for thinking that all of the necessary conditions for applying $P$ are satisfied.

The Evidentiary Principle defines a moral duty to ensure that one is epistemically justified in acting under the relevant moral principles. If one person $A$ takes aggressive action against another person $B$ without having sufficient evidence for believing the application-conditions of the relevant principle have been satisfied, $A$ has wronged $B$.

Accordingly, the victim of a digital attack can permissible adopt active response measures only if he or she has adequate reason to think the application-conditions of at least one of the relevant principles are satisfied. Under the Defense Principle, the victim 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 the attack from resulting in harm of some kind, and (3) such force is directed only at persons who are responsible for the attack. Under the Necessity Principle, the victim 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. If the victim lacks reason to think the application-conditions of both rules are satisfied and if these are the only relevant rules, then it would be wrong for him or her to adopt active response measures that infringe the rights of any innocent person.

\textsuperscript{52}This line of analysis, however, presupposes that the state is reasonably effective in protecting against such attacks. See the discussion below of this issue and how it bears on the ethics of active response.
necessary to achieve the greater moral good of preventing the damage caused by an attack, an aggressive response may result in unpredictable harms that outweigh the relevant moral goods. Machines can be linked via a network to one another in unpredictable ways, making it impossible to identify all the harmful effects of an aggressive response in advance. Ethically significant "collateral damage" can be ruled out reliably in only a small class of exceptional cases.

Indeed, a variety of intranational and international worst-case scenarios are unfortunately possible. 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 that 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 here is not that we have reason to think that these worst-case scenarios are very likely; rather, 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. Because the victim cannot reliably assess these probabilities, he or she lacks adequate reason to think that the application-conditions of the Necessity Principle are satisfied. Thus, under the Evidentiary Principle, the victim may not justifiably adopt aggressive measures under this principle.

Intermediate Responses

Intermediate active responses typically include exploratory tracebacks that attempt to identify culpable attackers by following attack paths in reverse through innocent agent machines (if any) to the ultimate source of the attack, as well as devices that allow entry into a remote system for the purpose of gathering information. Though such responses are neither intended nor reasonably likely to cause harm, they are ethically problematic insofar as they are invasive in the following sense: to the extent that the use of a traceback results in an unauthorized entry onto innocent agent machines, it would appear to constitute a trespass—something that is presumptively impermissible.

One might think that such trespasses can be justified under the Necessity Principle.26 To the extent that intermediate responses can be used reliably to identify the culpable source of a digital attack for the purpose of prosecuting the responsible parties,27 they function to secure the important moral good of restoring the public peace by bringing wrongdoers to justice—a good that seems important enough to justify comparatively minor trespasses onto the property of innocent persons.28

Unfortunately, it is frequently unclear whether intermediate responses are likely to succeed in identifying culpable parties. A sophisticated attacker can insulate himself or 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. However, the greater the number of "hops in the chain" between attacker and victim, the less likely that intermediate responses will succeed in identifying the culpable party. Indeed, it is fair to say that the likelihood of identifying the culpable parties in such attacks by intermediate responses is morally negligible.

This means that the expected moral value (i.e., the magnitude of the good multiplied by the probability of realizing it) to be achieved by using invasive tracebacks is a lower value than is desirable. In contrast, the expected moral disvalue (i.e., the magnitude of the bad multiplied by the probability of realizing it) is significant; although it may be difficult to precisely quantify the magnitude of the evil involved in a trespass, the probability of committing a trespass in using invasive tracebacks against any reasonably sophisticated attack will be close to 1. This militates against the claim that the good that will be achieved by using invasive tracebacks is significantly greater than the bad that will be done.

Moreover, the use of intermediate responses can also have significant collateral impacts that are undesirable from a moral point of view. The use of invasive tracebacks can result in damage to a variety of important trust relationships. A private firm that implements a traceback in an attack staged from the machines of other competing businesses can damage not only trust relationships between those businesses but could also precipitate a response that damages trust relationships between consumers and businesses, potentially resulting in economic losses that are passed on to the public in the form of lost jobs. Even worse, the use of invasive tracebacks by a private firm in response to an attack staged from machines used by state officials of another nation could result in an international incident that damages the relationship between those nations.

Again, the point here is not that we have reason to think that these worst-case scenarios are very likely; rather, a victim contemplating an intermediate response cannot reliably estimate the probabilities of such scenarios in such a short period of time. Because the victim cannot reliably assess these probabilities, there is inadequate reason to think that the application-conditions of the Necessity Principle have been satisfied. The Evidentiary Principle seems to preclude adopting intermediate responses in ordinary cases in which the victim lacks fairly detailed knowledge about the source and routing of an attack.

26 Because intermediate responses neither punish nor defend against attacks, the ethical principles allowing defense or punitive measures are irrelevant.

27 Not all intermediate responses are motivated by a desire to prosecute the wrongdoer. Many firms would prefer to avoid prosecuting intruders to escape the unfavorable publicity that might result from the disclosure of security breaches. The above reasoning would not justify intermediate responses in these cases.

28 It is the only way that a private security officer can apprehend a robbery suspect if he or she is justified in doing so under the Necessity Principle.
Thus, in most cases, private parties cannot justify adopt-
ing intermediate responses on the strength of the Neces-
sity Principle.

Nevertheless, it is important to emphasize that the
analysis here is limited to current traceback technolo-
gies with their limitations. Many researchers are mak-
ing considerable progress in improving the reliability
and efficacy of traceback technologies (see, for exam-
ple, http://footfall.csc.ncsu.edu, which documents some
intriguing advancements in these technologies.)

Indeed, one might reasonably expect that researchers
will eventually improve these technologies to the point
where they are sufficiently efficacious in identifying cul-
pable parties. Then, they could generally be justified un-
der the Necessity Principle as bringing about the greater
moral good of identifying culpable parties to an attack.
Thus, although presumptively unjustified under the Ne-
cessity Principle at this juncture, this may not be the case
for long.

Benign Responses

One might think that benign responses are ethically un-
problematic because, by definition, they affect only those
physical resources owned by the victim. According to this
line of argument, a person has a moral liberty to dispose
of property as he or she sees fit. Other things being equal,
the property owner has a liberty to make physical alter-
ations in his or her property; thus, for example, I have a
liberty to make my home safer by installing a new lock
in my door. It might be true that one's obligatios to, say,
support one's family preclude damaging one's own prop-
erty in circumstances where doing so renders one unable
to support one's family; however, these constraints are ex-
ceptional. Persons are at considerable liberty to use or
modify their property as they see fit.

This argument is problematic in a couple of ways. First,
the party implementing a benign measure may be a non-
owner who has authorized control over the owner's re-
sources. The scope of the owner's liberty to dispose of his
or her property does not extend in an unrestricted fashion
to agents of the owner. What the agent may permissibly
do depends on other factors including the terms of the agree-
ment he or she has with the owner.

Second, and more important, the scope of a person's
moral liberty with property is limited by the rights of other
persons. For example, the fact that an employer E owns
a workspace does not imply that E has a right to install
cameras in the bathroom to monitor employees. In this
case, E's liberty to dispose of E's property is outweighed
by the right of E's employees to privacy.

Accordingly, the mere fact that benign measures affect
only the property of the victim is not by itself sufficient
to imply that they are ethically justified; if there are other
persons, for example, who have privacy rights (such as
might be true of an ISP) that might be violated by benign
measures, then such measures are not clearly permissi-
ble. Indeed, users of a system might have a reasonable ex-
ppectation of privacy in the contents of their files or com-
munications that give rise to privacy rights that would
be violated even by benign measures. Such expectations
might arise, for example, among users of a university or
 corporate network.

For this reason, it is not possible to draw any general
conclusions about the permissibility of benign responses
to digital attacks. Because they will be impermissible in
cases where they violate the rights of third parties, it is
necessary in any given instance to determine whether
there are third parties who have rights that might be vio-
lated by the adoption of even benign responses.

The Inadequacy of Law Enforcement Efforts

There is, however, one powerful argument that can be
made in defense of the view that it is permissible for
private individuals to undertake active response. The ar-

gument 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 it
is providing minimally adequate protection against such
attacks. If (1) digital intrusions are resulting in signifi-
cant harm or injury of a kind that the state ought to pro-
tect against and (2) the state's protective efforts are inade-
quate, then private individuals, in this line of reasoning,
are entitled to adopt active response measures that con-
due to their own protection.

Both antecedent clauses appear to be satisfied. Depend-
going on the target and sophistication of the attack, an unau-
thorized digital intrusion can result in significant financial
losses to its victims. For example, an extended DoS at-
tack that effectively takes a major online retailer 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 the loss of value to shareholders and ultimately the 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 pur-
sue investigations in the vast majority of computer intru-
sions. Even when there are sufficient resources to justify
the state in intervening, the response is likely to come
long after the damage is done. That law enforcement has
just not been able to keep pace with the rapidly growing
problems posed by digital attackers is not a matter of con-
m treverety.

There are a variety of reasons for this inadequate re-
sponse. Most obviously, the availability of resources for
combating cybercrime is constrained by fiscal and polit-
ical realities; if the public is vehemently opposed to tax
increases that would increase the resources for investi-
gating cybercrime, then the growth of those resources
will not keep pace with an increasing rate of intrusions.
Equally important, there are special complexities involved
in investigating and prosecuting digital intrusions. First,
according to Mitchell and Banker (1998), investigation of
digital intrusions is resource-intensive: "[W]hereas a typi-
cal (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 job handling three or four
cases a month." Second, most sophisticated attacks pose
jurisdictional complexities that increase the expense of
law enforcement efforts because such attacks 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.

Though such considerations show that the growing problem associated with digital intrusions demands an effective response of some kind, they fall short of showing that it is permissible, as a general matter, for private parties to undertake intermediate or aggressive active response measures. The argument above assumes that private individuals can do what the state cannot—namely, protect themselves adequately from the threats posed by digital intrusion. That is to say, the argument assumes that private active response is likely to be efficacious in achieving legitimate objectives.

At this time, however, there is very little reason to think that this underlying assumption is correct. For starters, invasive intermediate measures intended to collect information are likely to succeed in identifying culpable parties, as noted above, in only direct attacks staged from the attacker's own computer, or by unskilled or careless attackers; such measures are not likely to succeed in identifying parties culpable for intrusions that are staged from innocent machines or where a high degree of preparation and sophistication are involved. Because 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 attacker and target, there is little reason to think that invasive investigatory measures are likely to achieve their legitimate 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 succeed in protecting 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 instead frequently result in escalating the attack; for this reason, such countermeasures are not likely to succeed in achieving legitimate objectives that are purely defensive in character. Further, aggressive countermeasures 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 the target will respond with aggressively punitive measures can simply interpose an additional layer of innocent machines to insulate himself from the target.

Yet, if the above argument fails to justify active response by private victims, it succeeds in showing that the problem of digital intrusions needs an effective coordinated solution of some kind—one that involves, at the very least, the sanction and cooperation of the state. One notable proposal deserves mention here. Mitchell and Banker (1998) have suggested a private-public solution that involves state licensing of security professionals who are trained in response to digital intrusions and who are authorized to do so subject to certain constraints. It is reasonable to hypothesize that an evaluation of such proposals will become the focus of normative research on active response in the near future.

THE LEGALITY OF ACTIVE RESPONSE

Although no Western nation has any statutes that explicitly address the legality of active response, there are a number of laws that potentially apply to it. This section discusses some of these laws and their potential application to the various levels of active responses. Nevertheless, it should be emphasized that the discussion in this section is tentative (and should not in any event be construed as authoritative legal advice). It is clear only that the legality of the various active responses remains unsettled at this time in every Western nation.

The United States

The federal law most likely to create liability for active response is Section 1030(a)(5) of the Computer Crime and Fraud Act (i.e., 18 U.S. Code Section 1030), which provides as follows:

Whoever . . . (A)(i) knowingly causes the transmission of a program, information, code, or command, and as a result of such conduct, intentionally causes damage without authorization, to a protected computer; (ii) intentionally accesses a protected computer without authorization, and as a result of such conduct, causes damage; or (iii) intentionally accesses a protected computer without authorization, and as a result of such conduct, causes damage; and (B) by conduct described in clauses (i), (ii), or (iii) of subparagraph (A), caused (or in the case of an attempted offense, would, if completed, have caused)—(I) loss to 1 or more persons during any 1-year period . . . aggregating at least $5,000 in value; (ii) the modification or impairment, or potential modification or impairment, of the examination, diagnosis, treatment, or care of 1 or more individuals; (iii) physical injury to any person; (iv) a threat to public health or safety; or (v) damage affecting a computer system used by or for a government entity in furtherance of the administration of justice, national defense, or national security . . . shall be punished as provided in subsection (c) of this section.

Subsection (c) authorizes fines and imprisonment of up to twenty years for specified violations of the quoted provision.

Although the relevant provisions apply only to "protected computers," the definition of that category is potentially broad. In particular, it includes any "computer . . . used in interstate or foreign commerce or communication." Construed literally, this provision would include any computer that has been used to send
an email from a person in one state to a person in another state or used to access any Web page that is published on a network in a different state from the user—which would seem to include every computer capable of being attacked. As the statute does not make exceptions for active response, persons adopting active responses to hacker attacks could potentially be prosecuted under the Act. For starters, it is highly likely that any computer being used in a hacker attack will satisfy the definition of a “protected computer,” if an active response measure results in statutorily sufficient damage, it could result in liability. It is true, of course, that the Act is most likely to apply to aggressive measures as these are intended to inflict damage or harm on the attacker. However, it could also apply to intermediate responses, such as invasive traceback, that deliberately trespass against external machines to identify attack paths. In the event that such measures proximately result in damage to those machines, they can give rise to liability under the Act.

Indeed, even benign measures might give rise to liability under certain circumstances. The mere fact that the user has property rights in a network does not, by itself, imply that the user is legally authorized to access computers on that network because computer users may have privacy rights that insulate their computers from certain kinds of access. To the extent that network owners access computers or files on those computers protected by such rights without appropriate authorization, they might be subject to liability for any benign measures that result in the right kinds of damage.28

Canada

Canada has several statutes potentially applicable to active response. For example, Section 342.1(1) of the Canadian Criminal Code provides:

> Every one who, fraudulently and without colour of right (a) obtains, directly or indirectly, any computer service, (b) by means of an electromagnetic, acoustic, mechanical or other device, intercepts or causes to be intercepted, directly or indirectly, any function of a computer system, (c) uses or causes to be used, directly or indirectly, a computer system with intent to commit an offence under paragraph (a) or (b) or an offence under section 430 in relation to data or a computer system, or (d) uses, possesses, traffics in or permits another person to have access to a computer password that would enable a person to commit an offence under paragraph (a), (b) or (c) is guilty of an indictable offence and liable to imprisonment for a term not exceeding ten years, or is guilty of an offence punishable on summary conviction.

Another statute that is potentially applicable to active response is Section 430 (11), which defines the crime of mischief as follows: “Every one commits mischief who willfully (a) destroys or alters data; (b) renders data meaningless, useless or ineffective; (c) obstructs, interrupts or interferes with the lawful use of data; or (d) obstructs, interrupts or interferes with any person in the lawful use of data or denies access to data to any person who is entitled to access thereto.”

Literally construed, the language of both sections seems applicable to the most aggressive active response measures. Section 430 requires that destruction, alteration, obstruction, or interruption of data be “willful” as is true of the most aggressive active response measures after all, such measures are, by definition, calculated to inflict harm or damage on those computers from which digital attacks are staged. Section 342.1 not only applies to these measures by incorporating the requirements of Section 430, but also by setting a less stringent standard for violation: merely obtaining unauthorized access (a necessary precondition for inflicting the sort of damage intended by aggressive active response) seems sufficient to subject a person to liability.

In contrast, only Section 342.1 seems applicable to intermediate responses, because intermediate responses are not intended to result in damage or destruction to targeted systems, Section 430 would not seem to apply. Section 342.1 is potentially applicable to a large range of intermediate responses precisely because there is no minimum damage requirement. In the absence of an applicable defense, unauthorized access of some kind seems sufficient to support liability for intermediate responses under Section 342.1.

The legality of benign responses under Section 342.1 turns on the same issues as discussed above in connection with the Computer Crimes and Fraud Act.

The European Union

On November 8, 2001, the Committee of Ministers of the Council formerly adopted the Convention on Cybercrime, which states guidelines for the various members of the European Union in formulating law regarding computer misuse (Council of Europe, 2001). Section 1 of Chapter II of the Convention states guidelines for formulating substantive criminal law as it pertains to unauthorized access to computers, unauthorized interception of data, data interference, system interference, misuse of computing devices, computer fraud, child pornography, and copyright infractions. Article 2 of the Convention is of particular relevance for our purposes as it defines the relevant guidelines for criminalizing unauthorized access of computer technologies. Article 2 provides as follows:

> Each Party shall adopt such legislative and other measures as may be necessary to establish as criminal offences under its domestic law, when committed intentionally, the access to the whole or any part of a computer system without right. A Party may require that the offence be committed by infringing security measures, with the intent of obtaining computer data or other dishonest intent, or in relation to a computer system that is connected to another computer system.
Insular as the defining characteristic of a hacker attack is the attempt to gain unauthorized access. Article 2 purports to guide the adoption of criminal laws regarding hacker attacks—and potentially active response. Notably, the Convention on Cybercrime is also motivated by a concern to address the problems that arise out of the transnational character of cyberspace: "Given the cross-border nature of information networks, a concerted international effort is needed to deal with such misuse." Chapter III defines the guidelines for international cooperation. Article 23 expresses the general tenor of the principles governing international cooperation: "The Parties shall co-operate with each other, in accordance with the provisions of this chapter, and through application of relevant international instruments on international cooperation in criminal matters, agreements arranged on the basis of uniform or reciprocal legislation, and domestic laws, to the widest extent possible for the purposes of investigations or proceedings concerning criminal offences related to computer systems and data, or for the collection of evidence in electronic form of a criminal offence." Remaining Articles define principles of extradition and other principles requiring mutual assistance among nations.

Legal Analogues of the Defense and Necessity Principles

The criminal law of most, if not all, Western nations incorporates principles that allow for the forceful defense of innocent persons and that allow for an otherwise wrongful act when necessary to secure a good that is significantly greater than the evil created by the commission of that act.

For example, the penal statutes of New York State contain analogues of both the Defense and Necessity Principles. Section 35.05 states the analogue of the Defense Principle: "A person may...use physical force upon another person when and to the extent he reasonably believes such to be necessary to defend himself or a third person from what he reasonably believes to be the use or imminent use of unlawful physical force by such other person."30 Section 35.05 states the analogue of the Necessity Principle: "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."31

These defenses, however, do not necessarily apply to the active response continuum. First, such statutes typically incorporate an evidentiary precondition for acting justifiably under the relevant defenses. Section 35.15 allows force only "when he reasonably believes" it is necessary. Similarly, Section 35.05 allows an otherwise wrongful act only when the moral goods "clearly outweigh" the moral evils. Even assuming that the courts in any particular jurisdiction are willing in principle to apply these sorts of provisions to active response contexts, the same sorts of defenses arise in the criminal context as in the ethical context.

Second, these statutes have been enacted with certain paradigmatic contexts in mind. They usually do not include active response measures. Consider, for example, the Model Penal Code’s comment on the necessity defense:

Under this section, property may be destroyed to prevent the spread of a fire. A speed limit may be violated in pursuing a suspected criminal. An ambulance may pass a traffic light. Mountain climbers lost in a storm may take refuge in a house or may appropriate provisions. Cargo may be jettisoned or an embargo violated to preserve the vessel. An alien may violate a curfew in order to reach an air raid shelter. A druggist may dispense a drug without the requisite prescription to alleviate grave distress in an emergency.32

As is readily evident, these examples do not anticipate the application of the necessity defense to digital contexts. For such reasons, it is simply not clear that courts are willing to apply these defenses to the digital context.

CONCLUSIONS

The active response continuum comprises a variety of noncooperative measures that are intended to respond to a digital intrusion in ways that have a causal impact on remote machines. Such measures include acts that would fairly be characterized as hacking in circumstances in which they were unprovoked by an intrusion. They range from more benign measures intended to inflict no damage to highly aggressive measures intended to inflict the same kind of damage on the attacker that he or she is attempting to inflict on the victim. These tactics include sniffing, scanning, traceback, corruption of data, remote exploitation, and denial of service attacks.

Because of their invasive quality, active response measures raise a variety of normative issues. Denial of service counterattacks, for example, raise prudential, ethical, and legal issues. Responding to a DoS attack with proportional force might actually result in increased harm to the victim insofar as it evokes an escalation of the attack; it might thus adversely affect the victim’s own interests. Furthermore, if the attack is distributed, the counterattack will necessarily be directed at innocent agent machines, raising serious ethical and legal issues. For all these reasons, such aggressive measures are ill-advised.

Nevertheless, the growing frequency of hacker attacks and the increasing inability of law enforcement agencies to respond adequately suggest the need for a coordinated solution involving both public and private elements. The key to keeping owners/operators from resorting to noncooperative, invasive active response measures might be for

---

governments to provide resources that would allow sites to enlist the support of trained and trusted regional incident response teams, thereby decreasing the differential in skills/resources/information that is driving the private sector to resort to active response.

There is now an opportunity for government policy-makers, the military, the law enforcement community, and the private sector to work together and develop effective cooperative relationships, operational resources and training, and a clear and streamlined legal framework that will level the playing field between attackers and defenders. The hope is that this will happen before a massive cyberattack occurs.

GLOSSARY

Active Response a variety of reactive, noncooperative responses to a digital intrusion that are typically calculated to affect remote systems and are intended to investigate, defend, repel, or punish the intrusion. Such measures range from benign measures that implicate the legitimate interests of innocent persons without affecting remote systems to aggressive measures that are intended to inflict harm or damage on the intended targets. Also referred to as "hacking back" and "counterhacking."

Agent Machines machines belonging to innocent parties that are compromised by an attacker and used to stage a digital attack or intrusion of some kind.

Cybercrime criminal activity that involves unauthorized use of computer technology in an essential way.

Cyberterrorism hacking activity that attempts to harm innocent persons and thereby create a general sense of fear or terror among the general population for the purpose of achieving a political agenda.

Denial of Service a digital attack that is calculated to shut down a Web site, server, or network, usually by overwhelming it with sham requests for information.

Digital Intrusion an act intended to gain unauthorized access to the digital contents (e.g., files or programs) of another person. Such access can be for comparatively benign purposes (e.g., merely to look at files) or can be for malicious purposes (e.g., to destroy files).

Hackers persons who attempt to gain unauthorized entry to network servers or other computers. Hacking is usually distinguished from "cracking" in that the latter, unlike hacking activity, is intended to cause harm to innocent persons.

Honeypots computer resources installed for the purpose of being compromised that are used as decoys that permit users to identify malicious activity.

Intrusion Response measures adopted by the victim of a digital intrusion intended to investigate, repel, or punish the intrusion.

Kantianism ethical theory that assesses action according to whether the underlying principle can consistently be universalized; acts, in this view, are intrinsically right or wrong.

Order of Volatility ranking of sources of digital information according to stability and durability.

Siphers programs that are designed to monitor network traffic on local area network segments.

Tracebacks programs designed to follow the path of an ongoing attack in reverse in order to identify its ultimate and hence culpable source.

Utilitarianism a consequentialist moral theory that holds that the goodness or badness of an action is determined entirely by its consequences on well-being, happiness, the number of preferences satisfied, or pleasure in the community; acts, in this view, are right or wrong in virtue of extrinsic characteristics (i.e., their effects), and not in virtue of intrinsic characteristics.

CROSS REFERENCES

See Hackers, Crackers and Computer Criminals; Intrusion Detection Systems Basics; Legal, Social and Ethical Issues of the Internet; Network Attacks.

REFERENCES


FURTHER READING