Framework and Principles for Active Cyber Defense[1]

Dorothy E. Denning
Naval Postgraduate School

December 2013

Abstract

This essay offers a broad view of active defense derived from the concept of active air and missile defense. This view admits a range of cyber defenses, many of which are widely deployed and considered essential in today's threat environment. Instead of equating active defense to hacking back, this wider interpretation lends itself to distinguishing different types of active defense and the legal and ethical issues they raise. The essay will review the concepts of active and passive air and missile defenses, apply them to cyberspace, describe a framework for distinguishing different types of active cyber defense, and finally suggest legal and ethical principles for conducting active cyber defense.

Keywords

Cyber security, cyber ethics, active defense, hacking back, air and missile defense.

Introduction

The concept of active cyber defense has raised red flags within the computer security community. Gary McGraw, Chief Technology Officer of Cigital, for example, has called it "irresponsible" and a "recipe for disaster," adding, "The last thing we need in computer security is a bunch of vigilante yoo-hoos and lynch mobs." (McGraw 2013) His remarks are based largely on a concept of active defense based on "hacking back" or "attacking the attacker," with the possibility of harming innocent persons in the process. Surely, if this is what active defense is all about, then it *should* give us pause.

This essay offers a broader view of active defense derived from the concept of active air and missile defense used by the US Department of Defense. This view admits a range of cyber defenses, many of which are widely deployed and considered essential in today's threat environment. Instead of equating active defense to hacking back, this wider interpretation lends itself to distinguishing different types of active defense and the legal and ethical issues they raise. The essay will review the concepts of active and passive air and missile defenses, apply them to cyberspace, describe a framework for distinguishing different types of active cyber defense, and finally suggest legal and ethical principles for conducting active cyber defense. It draws on work done in collaboration with colleague

[1] The views expressed in this document are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

and ethicist Bradley Strawser in the Defense Analysis Department at the Naval Postgraduate School (Denning and Strawser 2013).

Active and Passive Air and Missile Defense

US military doctrine distinguishes between active and passive air defenses. It defines *Active Air and Missile Defense (AMD)* as: **"**direct defensive action taken to destroy, nullify, or reduce the effectiveness of air and missile threats against friendly forces and assets." Active AMD is said to include "the use of aircraft, AD [air defense] weapons, missile defense weapons, electronic warfare (EW), multiple sensors, and other available weapons/capabilities" (JP 3-01 2012). It characterizes such actions as shooting down or diverting incoming missiles and jamming hostile radar or communications.

An example of an active air and missile defense system is the Patriot surface-to-air missile system, which uses an advanced aerial interceptor missile and high performance radar system to detect and shoot down hostile aircraft and tactical ballistic missiles (Patriot 2012). Patriots were first deployed in Operation Desert Storm in 1991 to counter Iraqi Scud missiles. Israel's Iron Dome anti-rocket interceptor system has a similar objective of defending against incoming air threats. According to reports, the system intercepted more than 300 rockets fired by Hamas from Gaza into Israel during the November 2012 conflict, with a success rate of 80 to 90 percent (Kershner 2012). At the time, Israel was also under cyber assault, and Prime Minister Benjamin Netanyahu said that the country needed to develop a cyber defense system similar to Iron Dome (Ackerman and Ramadan 2012).

Another example of an active air defense system is the US's Operation Noble Eagle (Air Force 2012). Launched minutes after the first aircraft was hijacked the morning of September 11, 2001, the operation has become a major element of homeland air defense through its combat air patrols, air cover support for special events, and sorties in response to possible air threats. Although Noble Eagle pilots can potentially shoot down hostile aircraft, so far none have done so. However, they have intercepted and escorted numerous planes to airfields over the years.

In contrast to active defense, *Passive Air and Missile Defense* is defined as: "all measures, other than active AMD, taken to minimize the effectiveness of hostile air and missile threats against friendly forces and assets," noting that "these measures include detection, warning, camouflage, concealment, deception, dispersion, and the use of protective construction. Passive AMD improves survivability by reducing the likelihood of detection and targeting of friendly assets and thereby minimizing the potential effects of adversary reconnaissance, surveillance, and attack." (JP 3-01 2012) It includes such actions as concealing aircraft with stealth technology. It covers monitoring the airspace for adversary aircraft and missiles, but not actions that destroy or divert them.

Active and Passive Cyber Defense

The definitions of active and passive air defense can be applied to the cyber domain by replacing the term "air and missile" with "cyber." This gives: *Active Cyber Defense* is direct defensive action taken to destroy, nullify, or reduce the effectiveness of cyber threats against friendly forces and assets. *Passive Cyber Defense* is all measures, other than active cyber defense, taken to minimize the effectiveness of cyber threats against friendly forces and assets. Whereas active defenses are direct actions taken against specific threats, passive defenses focus more on making cyber assets more resilient to attack.

Many popular security controls employ active cyber defenses. Access controls block users from accessing unauthorized files and other resources. Passwords and other user authentication mechanisms block login attempts from adversaries spoofing as legitimate users. Anti-malware systems, intrusion prevention systems (IPSs), and firewalls block malicious software and packets matching threat signatures or exhibiting anomalous behavior. Honeypots lure or deflect attacks into isolated systems where they can be monitored and kept away from production systems. All of these controls are analogous to air and missile defenses that shoot down or deflect incoming missiles and rockets. Active cyber defenses also include operations against systems owned or used by an attacker, including counter-attacks. These are more analogous to air defense operations that attack the air or ground platforms used by the adversary to launch missiles.

Passive cyber defenses include cryptography and steganography (analogous to the use of camouflage and stealth aircraft), security engineering and verification, configuration monitoring and management, vulnerability assessment and mitigation, risk assessment, backup and recovery of lost data, and education and training of users. They also include mechanisms to log and monitor network and host activity (analogous to air monitoring). Intrusion detection systems (IDSs) are essentially passive, but become active when they incorporate elements to abort detected threats, morphing into IPSs.

A Framework for Active Cyber Defenses

Active cyber defenses can be characterized by four features: scope of effects, degree of cooperation, types of effects, and degree of automation. Together, they place active cyber defenses in a four-dimensional space and provide a framework for distinguishing different types of active cyber defenses and analyzing the ethical issues they raise.

Scope of effects. This feature distinguishes between *internal* defenses, whose effects are limited to the network being defended, and external defenses, whose effects go beyond the network. An internal cyber defense is akin to an air defense system that takes actions against an incoming missile or hostile aircraft after it has entered a country's airspace, while an external cyber defense is like an air defense system that takes action in someone else's airspace. Most cyber security controls such as access controls and IPSs are internal. An example of an active defense with external effects is a botnet takedown that involves taking over the IP addresses and domain names used for command and control (C2).

Degree of cooperation. This feature distinguishes between active defenses that are *cooperative*, meaning that action is one that is performed against a system with the knowledge and consent of the system owner, from those that are *non-cooperative*, meaning it is not. Using the air defense analogy, a cooperative cyber defense with external effects is like an air defense system that shoots down missiles in the airspace of an ally that has requested help, while a non-cooperative cyber defense is like an air defense system that shoots them down in the adversary's own airspace.

Cyber defenses that involve hacking back or attacking the attacker fall in the category of non-cooperative, external operations. A good example is an operation performed by the Georgian government against a Russian-based hacker who had waged a persistent, month-long campaign to steal confidential data from Georgian systems. Using a "water-hole" attack, the hacker had managed to infect Georgian computers with malware that exfiltrated files matching certain criteria to a drop site belonging to the hacker. To counter the hacker, the government planted a decoy ZIP archive on one of its infected machines, which the malware dutifully exfiltrated to the drop site. Once the hacker downloaded and opened the archive, it unleashed spyware that passed data from the hacker's machine back to the Georgian government, including a photo of the hacker taken by his own webcam (Kirk 2012). It is worth noting, however, that whereas the operation to plant spyware on the hacker's system was non-cooperative, it was the hacker's own actions and code that caused his system to be infected. The Georgian government did not directly hack his machine or any of the servers he used.

Types of effects. This feature distinguishes four types of effects. The first, *sharing*, refers to actions that distribute threat information such as the IP addresses of attacking computers or the signatures of attack packets to other parties. Sharing is involved when anti-malware vendors ship out new signatures to their customers or victims report the domain names or IP addresses of malicious sites.

The second type of effect, *collecting*, is one that takes actions to acquire more information about the threat, for example, by activating or deploying additional sensors or by serving a court order or subpoena against the source or an ISP likely to have relevant information. When the Coreflood botnet was taken down, for example, its attacker-controlled C2 servers were effectively replaced with C2 servers operated by the non-profit Internet Systems Consortium in collaboration with the federal government. The servers were set up to collect the IP addresses of the bots when they checked in for instructions. These addresses were then shared with the FBI, which in turn shared them with their associated ISPs so that the victims could be notified. The spyware used in the Georgian hacking case described above also illustrates collecting (Zetter 2011a, 2011b; Higgins 2011).

The third type of effect, *blocking*, is one that blocks activity deemed hostile, for example, traffic from a particular IP address or execution of a particular program. The Coreflood takedown had the effect of breaking the communication channel from the persons who had been operating the botnet to the bots. As a result, they could no longer send

commands to the bots. Further, when the bots contacted the C2 servers, they were given a "stop" command, effectively blocking any further bot activity, but without interfering with other activity on the infected computers (Zetter 2011a, 2011b; Higgins 2011). In the Georgian case, most connections to the drop servers were blocked in order to prevent further exfiltration of sensitive data (Kirk 2012). Access controls, firewalls, anti-malware controls, and IPSs also illustrate blocking.

Finally, the effects are said to be *preemptive* if they neutralize or eliminate a source used in the attacks, for example, by seizing the computer of a person initiating attacks or by taking down the command and control servers for a botnet. In the Coreflood takedown, the hostile C2 servers were effectively put out of commission and the bots neutralized. With further action on the part of victims, the malware could also be removed.

Using the air defense analogy, the cyber defense of sharing is like a missile defense system that reports new missile threats to allies so that they can counter them. The cyber defense of collecting is like a missile defense system that activates additional radars or other sensors in response to an increased threat level, or that sends out sorties to investigate suspicious aircraft. The cyber defense of blocking is akin to a missile defense system that shoots down or deflects incoming missiles or jams their radars and seekers, thereby preventing them from hitting their targets. Finally, the cyber defense of preemption is like launching an offensive strike against the air or ground platform launching the missiles.

Degree of automation. This feature pertains to the degree of human involvement. An active defense is said to be *automatic* if no human intervention is required and *manual* if key steps require the initiation or affirmative action of humans. Most anti-malware and intrusion prevention systems have both manual and automated components. Humans determine what goes into the signature database, and they install and configure the security software. The processes of signature distribution, malicious code and packet detection, and initial response are automated, but humans may be involved in determining the final response.

In the Coreflood takedown, the transmission of the stop commands was fully automated through the C2 servers. However, humans played an important role in planning and decision making, analyzing the botnet code and the effects of issuing a stop command, acquisition of the restraining order, and swapping out of the C2 servers. Thus, the entire operation had both manual and automatic aspects. In the Georgian case, much of the investigation involved manual work, including analyzing the code, determining what the hacker was looking for, and setting up the bait with the spyware. But the key element in the outing, namely the operation of the spyware, was automated. Once the hacker downloaded the ZIP archive, it did the rest.

Applying the air defense analogy, an automatic cyber defense is like an anti-missile system that automatically shoots down anything meeting the preset criteria for being a hostile aircraft or incoming missile, whereas a manual cyber defense is more like Operation Noble Eagle where humans play a critical role both in recognizing and

responding to suspicious activity in US airspace. However, even anti-missile systems require humans to specify their settings and where and when they are deployed.

Ethical and Legal Principles for Active Cyber Defense

Active cyber defenses should be employed only when doing so is ethical and legal. The following six principles aim to promote that: authority, third party immunity, necessity, proportionality, human involvement, and civil liberties. The list is not prioritized, as all are important. It is derived from our earlier work, where we examine in greater depth the ethical issues raised along each of the four dimensions (Denning and Strawser), as well as from principles pertaining to state actors under the law of armed conflict and to non-state actors under domestic laws of self-defense.

Authority. Active cyber defenses should be conducted only with authorities granted by laws, contracts, and policies. For defenses that are internal only, such as the deployment of a firewall or IPS, network administrators generally have the necessary authorities. For defenses that are external but cooperative, such as the sharing of threat information, the authorities are somewhat more limited. For example, whereas security companies can send signature updates to their customers or manage firewalls and IPSs on their customers' networks, companies are reluctant to share certain threat information with each other or with the government lest they run up against anti-trust laws or find themselves subject to liabilities.

The issue of authority becomes especially problematic when defenses produce non-cooperative, external effects, say shutting down a botnet's C2 servers. In those cases additional authorities may be needed from the government, such as a court order. The execution of some active defenses may even be restricted to government agencies with appropriate authorities and responsibilities.

In general, governments, particularly their law enforcement, national security, and homeland defense arms, have greater authority than the private sector to conduct external, non-cooperative active defenses. In the area of active air and missile defense, the military has almost exclusive authority. Private sector entities are neither authorized nor expected to operate anti-missile systems, even to protect their own property. Cyber is different, as every network owner is expected to provide strong defenses, including active defenses, to defend their network. Still, whether they can employ certain active defenses, particularly when they involve a counter-attack against an adversary network, is controversial. For example, suppose the Georgian operation had been conducted by a company rather than the government. Even though the spyware in this case would have been planted on the company's own network, because it made its way to the hacker's system, it effectively would have given the company access to information on the hacker's computer, something the hacker himself surely would not have authorized. For an interesting and informative debate on the legality of such private sector counter-hacks under US law, see (Steptoe 2012).

Third party immunity. Active cyber defenses should not intentionally harm third parties. This includes any third party systems compromised by the attacker and used in the attacks. For example, an operation to shut down a botnet should not harm victim machines on the botnet, even if they have been spewing packets as part of a DDoS operation. Indeed, most if not all of the botnet takedowns, including the Coreflood takedown, have avoided harming the computers hosting the bots.

This principle is derived from the related principles of distinction and noncombatant immunity in just war theory. Noncombatants, including their property, are to be distinguished from military forces and not directly targeted. Still, just war theory recognizes that some collateral damage may be morally permissible when the actions are necessary and the harm proportionate to the gains. The next two principles put these in the context of active cyber defense.

Necessity. Active cyber defenses should not be deployed unless they are necessary to mitigate the threat. This principle applies especially to operations that affect third parties: they should not be attacked or harmed in any way unless doing so is essential. Had the Coreflood takedown wiped or even disabled the computers running the bots, for example, the operation would have been morally impermissible under this principle as damaging these machines was unnecessary to neutralize the botnet.

The principle of necessity also applies to operations that target the source of an attack, implying they should not be conducted for the sole purpose of retaliation or retribution. Rather, they should be conducted only as necessary to mitigate or defend against a specific threat and without causing gratuitous harm, even to the attacker. Neither the Coreflood takedown nor the Georgian operation against the Russian hacker were retaliatory in nature.

Proportionality. Active cyber defenses should not be deployed unless the harm incurred is proportionate to the benefits gained. To illustrate, suppose a compromised server has been deployed in a major DoS attack against a bank. After detecting the heavy stream of attack packets, the bank's ISP starts blocking all traffic from the attacking server going through its network. In the process, the compromised server may be prevented from sending legitimate traffic as well. Because of the potential harm this could cause, the operation is morally permissible only if that harm is determined to be proportionate to the benefits gained.

The reason for not precluding all harm to third parties is nicely illustrated by Iron Dome, where the act of shooting down a rocket can cause some damage from the fallout. However, Israel launches their counterstrikes primarily at rockets aimed at densely populated urban areas (Kershner 2012), so the relatively small amount of harm that might result from any fallout is more than compensated by the greater harm that would be caused if the rockets hit their intended targets.

In the domain of cyber, it is incumbent on those applying active defenses to know what effects they may have, especially to third parties. Without that knowledge, the principle

of proportionality cannot be reliably applied to determine whether an active defense is morally permissible. For example, suppose the server in the above example were used to support life-critical medical equipment at remote locations. Then an operation that blocked legitimate traffic from the server could potentially have life or death consequences, causing disproportionate harm. While it may not be possible to anticipate all of the effects of an operation employing active defenses, such operations should be conducted only when there is good reason to believe they will be proportionate. In the above scenario, absent more information about the server, a better moral choice might be to block only the attack traffic between the server and the bank.

Because cyber attacks can be difficult to attribute, active defenses that target a presumed source must be applied carefully, especially if they can cause damage. In general, active defenses that involve blocking or preemption are the most problematic, as they affect the availability and integrity of data and systems.

Human involvement. Active cyber defenses should employ humans at some stage. Indeed, even automated active cyber defenses, including user access controls, firewalls, anti-malware controls, and IPSs, depend on humans to make or confirm their settings. Humans have also played an essential role in botnet takedowns, including the Coreflood takedown, and were heavily involved in the Georgian operation.

At the speed of cyber, placing humans in the loop at every step is neither practical nor desirable. For example, imagine an IPS that had to check with a human operator before blocking a packet that matched an attack signature.

Civil liberties. Active cyber defenses should respect the civil liberties of all persons affected, including their rights of privacy, free speech, and association. This principle applies to users of the defending network as well as third parties and suspects. The right to privacy especially relates to active defenses that trigger the additional collection or sharing of information that contain personal information. In the Coreflood takedown, the replacement C2 servers were set up to collect and then share with ISPs the IP addresses of the computers running the bots in order that the computer owners could later clean their machines of the malware. The operation did not gather any information from the computers or post the IP addresses on a public site, so the effects on privacy were minimal.

The right to free speech and association applies especially to active defenses that involve blocking or preemption, as such operations constrain the flow of information. For example, if an organization blocks access to a legitimate site suspected of hosting malware, their own users will be prevented from accessing the site. While this may be justified in light of the threat, the implications of blocking access to a particular site should at least be considered.

Conclusions

Ethical and legal issues relating to active cyber defense have been discussed and debated for well over a decade now. While it is beyond the scope of this essay to review the literature in this area, the interested reader will find links to papers, presentations, and other materials on active defense on David Dittrich's web page on active defense (Dittrich 2013). Dittrich and colleague Kenneth Himma have examined the legal and ethical issues of active cyber defense along a continuum of non-cooperative tactics ranging from benign (inflicting no damage) to intermediate and then aggressive (inflicting damage comparable to that received) use of force (Dittrich and Himma 2005).

The framework presented in this essay views active cyber defense not as a continuum defined by aggressiveness, but rather as a multi-dimensional space characterized by four features or dimensions: scope of effects (internal and external), degree of cooperation (cooperative and non-cooperative), type of effects (collecting, sharing, blocking, and preemptive), and degree of automation (automatic and manual). The framework builds on concepts from air and missile defense, and examples in that domain are used to illustrate tactics, concepts, and ethical issues relating to cyber defense. The framework is also accompanied by a set of six principles for analyzing the ethics and legality of active cyber defenses: authority, third party immunity, necessity, proportionality, human involvement, and civil liberties. The framework and principles allow us to go beyond simply viewing all active cyber defenses as irresponsible, illegal, or dangerous, and instead draw on the distinctions provided to evaluate particular defenses.

REFERENCES

Ackerman, G. and Ramadan, S. A. (2012) 'Israel Wages Cyber War With Hamas as Civilians Take Up Computers,' *Bloomberg*, November 19. http://www.bloomberg.com/news/2012-11-19/israel-wages-cyber-war-with-hamas-as-civilians-take-up-computers.html (accessed November 26, 2012).

Air Force. (2012) Operation Noble Eagle, Air Force Historical Studies Office, Posted September 6. http://www.afhso.af.mil/topics/factsheets/factsheet.asp?id=18593 (accessed November 6, 2012).

Denning, D. E. and Strawser, B. J. (2013) 'Active Cyber Defense: Applying Air Defense to the Cyber Domain,' presented at the Cyber Analogies Seminar, May 3.

Dittrich, D. (2013) http://staff.washington.edu/dittrich/activedefense.html (accessed October 16, 2013).

Dittrich, D. and Himma, K. E. (2005) 'Active Response to Computer Intrusions,' *The Handbook of Information Security* (Bidgoli, H. ed.), John Wiley & Sons.

Higgins, K. J. (2011) 'Coreflood Botnet An Attractive Target For Takedown For Many Reasons,' *Dark Reading*, April 14. http://www.darkreading.com/database-security/167901020/security/client-security/229401635/coreflood-botnet-an-attractive-target-for-takedown-for-many-reasons.html (accessed November 27, 2011).

JP 3-01 (2012) 'Countering Air and Missile Threats,' Joint Publication 3-01, March 23.

Kershner, I. (2012) 'Israeli Iron Dome Stops a Rocket With a Rocket,' *The New York Times*, November 18. http://www.nytimes.com/2012/11/19/world/middleeast/israeli-iron-dome-stops-a-rocket-with-a-rocket.html?_r=0 (accessed November 19, 2012).

Kirk, J. (2012) 'Irked By Cyberspying, Georgia Outs Russia-Based Hacker – With Photos,' *Network World*, October 30. http://www.networkworld.com/news/2012/103012-irked-by-cyberspying-georgia-outs-263790.html (accessed November 27, 2012).

McGraw, G. (2013) '"Active Defense" is Irresponsible,' Cigital blog, February 14. http://www.cigital.com/justice-league-blog/2013/02/14/active-defense-is-irresponsible/ (accessed October 8, 2013).

Patriot. 'MIM-104 Patriot,' Wikipedia. http://en.wikipedia.org/wiki/MIM-104_Patriot (accessed November 6, 2012).

Steptoe. (2012) 'The Hackback Debate,' Steptoe Cyberblog, November 2. http://www.steptoecyberblog.com/2012/11/02/the-hackback-debate/ (accessed November 29).

Zetter, K. (2011a) 'With Court Order, FBI Hijacks Coreflood Botnet, Sends Kill Signal," *Wired*, April 13. http://www.wired.com/threatlevel/2011/04/coreflood/ (accessed November 27, 2012).

Zetter, K. (2011b) 'FBI vs. Coreflood Botnet: Round 1 Goes to the Feds,' *Wired*, April 26. http://www.wired.com/threatlevel/2011/04/coreflood_results/ (accessed November 27, 2012).