CERIAS Tech Report 2016-01

The Ethics of Hacking Back

by Corey T. Holzer, James E. Lerums
Center for Education and Research
Information Assurance and Security
Purdue University, West Lafayette, IN 47907-2086
The Ethics of Hacking Back

Corey T. Holzer, James E. Lerums
CERIAS
Purdue University
West Lafayette IN 47907
cholzer@purdue.edu, jlerums@purdue.edu

Abstract—Cyber breaches are increasing in frequency and scope on a regular basis. The targeted systems include both commercial and governmental networks. As the threat of these breaches rises, the public sector and private industry seek solutions that stop to the ones responsible for the attacks. While all would agree that organizations have the right to protect their networks from these cyber-attacks, the options for defending networks are not quite as clear. Few would question that a passive defense (i.e. the filtering of traffic, rejecting packets based on the source, etc.) is well within the realm of options open to a defender. What active defensive measures are ethically available to the defenders when passive options fail to stop a persistent threat is not as clear. This paper outlines the two (law enforcement and military) ethical frameworks commonly applied by cyber security professionals when considering the option of a cyber counter-offensive or “hacking back.” This examination includes current applicable literature in the fields of information security, international law, and information assurance ethics.

Keywords—Active Cyber Defense, Cybersecurity, Information Assurance Ethics, Laws of Armed Conflict, Law Enforcement

I. INTRODUCTION

The reporting of cyber breaches is increasing in frequency and scope on a regular basis and the systems being targeted include both commercial and government networks. As the threat that these breaches rise private industry and the public sector both seek solutions that will put a stop to those responsible for the attacks. While all would agree that organizations have the right to protect their networks from these cyber-attacks, the options for what includes legal and moral defense of one’s network is not quite as clear. Few would question that a passive defense (i.e. the filtering of traffic, rejecting packets based on the source, etc.) are well within the realm of options open to a defender. However, when these options are not enough to stop a persistent threat is an active defense, or counter-attack an option and, if so, under what circumstances is it ethical for a government organization able to employ such an option.

There are two overarching schools of thought or frameworks that information security experts, legal experts, and academics subscribe to when analyzing illegal or unethical cyber activity. The first describes these actions as criminal acts [1]. The second group holds that, depending on the nature of the act and its potential results, these acts should be considered acts of aggression and that a military style response is acceptable [2].

Our goal with this paper is to examine the ethicality of a cyber counter attack or what is commonly referred to as “hacking back” against cyber attackers. Before examining these two frameworks and applying them to the ethics of “hacking back” we will scope our work. Next we will review the frameworks for both the law enforcement and military approaches.

The authors of this work recognize that, as designed, the Laws of Armed Conflict are only legally binding to nation-state actors. However, our examination is intended to look at the ethical value of such a framework and how it can be employed in active cyber defense by Computer Network Defenders regardless of whether or not they are state actors. In short, there are ethical lessons to be learned through the study of this framework.

II. SCOPING OUR EXAMINATION

As we proceed with our examination of the ethics of hacking back, we will outline the frameworks for each of these schools of thinking and apply their perspective to the ethical appropriateness of the response of what we define as “hacking back.” First, we will outline the framework and rules that each school of thought expounds by a review of existing literature. With an understanding of each framework established, we will analyze scenarios using these frameworks with the goal of reaching an ethical conclusion regarding the response of those designated as the defenders of the network being attacked.

Regardless of the framework applied one of the greatest challenges in analyzing any cyber-attack is the matter of attribution. In this context, attribution is the determination of what party is responsible for a given cyber act [1], [3]–[5]. Experts have analyzed this issue extensively as a legal and ethical challenge and a fundamental challenge to any response to a cyber-attack. For our purposes in this endeavor, we will assume that attribution is certain and that our defenders can act with the knowledge that they know who is responsible for the attack against their network. Our intent is not to downplay the importance of the attribution dilemma but, instead, to address the ethical question surrounding the response of hacking back.

III. FRAMEWORK FOR THE LAW ENFORCEMENT APPROACH

The world experienced at the end of 2015 the largest data breach ever publicly reported (191 million records were exposed). Combined with eight other mega-breaches reported (defined as a breach of more than 10 million records) made 2015 a record-setting year with a twenty three percent increase in exposed identities totaling 429 million [6]. Breaches with a financial motive (greater than 75%) dominate everything else to include espionage and fun [7].
The increase of cyber-attacks fuels the perception that the government is unable to stem lawlessness in cyberspace. Today, it is not uncommon when major US companies find themselves as cyber victims they can usually only hunker down, endure the bad publicity and strengthen their defenses in hopes of deflecting future attacks [8].

Consequently, the concept of “hacking back” as a means to identify hackers and/or destroy stolen data is a topic being considered by company officials whereas it was once considered too reckless to suggest. The mere mention of “hacking back” within cybersecurity circles prompts warnings about the risks beginning with the illegality, and ending with the risk of triggering a full scale cyberwar with collateral damage across the Internet.

However, concerns over cyber breaches against the intellectual property of U.S. Corporations and espionage associated with breaches of U.S. government networks originating from China alone, prompted the U.S.-China Economic and Security Review Commission to make the following recommendation to Congress in its November 2015 report:

“Congress assess the coverage of U.S. law to determine whether U.S.-based companies that have been hacked should be allowed to engage in counter intrusions for the purpose of recovering, erasing, or altering stolen data in offending computer networks. In addition, Congress should study the feasibility of a foreign intelligence cyber court to hear evidence from U.S. victims of cyber-attacks and decide whether the U.S. government might undertake counter intrusions on a victim’s behalf [9, p. 564].”

Next, we will review the applicable U.S. Law and International Laws and Treatises which directly impact the private sector’s ability to respond to a cybersecurity breach. We will follow it with a review of non-binding standards and guidelines being developed in the International arena. With the laws and standards outlined, we will transition to discussing real world hack back example.

A. Applicable U.S. Law

Current United States law authorizes companies to deploy cybersecurity countermeasures on their own networks and systems against malware, and it criminalizes computers attacks on others (including hacking back) [10].

1) The Wiretap Act (1988): Authorizes providers of electronic communications service to intercept, disclose, or use communications passing through its network while engaging in any activity that is necessary incident to the protection of its rights and property (18 USC 2511(2)(a)(ii)). This includes the authority to use devices and procedures to intercept or redirect communications in order to protect their networks and the transiting data.

2) The Computer Fraud and Abuse Act (1986): The Federal anti-hacking law that subjects to criminal and civil liability anyone who intentionally accesses another person’s computer without authorization, and as a result of such conduct, recklessly causes damage (18 USC 1030(a)(5)(B)). In other word in the United States entering another person’s or company’s network in the course of attempting to identify hackers or destroy data they have stolen without permission violates the Computer Fraud and Abuse Act [8].

3) Cybersecurity Information Sharing Act of 2015: Authorizes operation of “defensive measures” (i.e. hack-back countermeasures) on one’s own network (or a network where written authorization has been granted). Prohibited are “defensive measures” that provides unauthorized access to, or substantially harms an information system or information stored on, processed by, or transiting information systems not belonging to the defender (or authorized in writing to defend) [11].

B. Applicable International Laws and Treatises

1) Charter of the United Nations, Article 2(4) (1945): Prohibits the “use of force” against the territorial integrity or political independence of another State. “Use of force” doesn’t have to be an “armed attack” and may include a cyber-attack based on a case by case assessments after considering factors such as severity, immediacy of effect, invasiveness, military character, and so forth [12].

2) Charter of the United Nations, Article 51 (1945): States have an inherent right of collective or individual self-defense in the case of an armed attack. Article 51 generally doesn’t apply to a cyber-attack unless it causes substantial injury or physical damage [12]. For example, a cyber-attack could be considered an armed attack if it causes an oil refinery plant to explode (as if detonated with explosives) resulting in a substantial number of fatalities and property damage.

3) International Law of Countermeasures: In general, international law supports regulating cyber space as an economic and communications sphere and contains coercive means of responding lawfully to cyber provocations of all types. The same sort of coercive measures that are lawful to use against economic wrongdoing will generally be lawful to use in the case of a cyber-attack. In the economic sphere, responses to violations tend to be known as countermeasures, and do not involve the use of military force [2].

Countermeasures are the mechanisms through which international law allows parties to carry out self-help, coercive enforcements of their rights. In the absence of an international level central police and compulsory courts, “self-help” plays a large role in international law enforcement. The International Court of Justice laid out the following four elements of a lawful countermeasure:

- The countermeasure must be taken in response to a previous international wrongful act of another State and must be directed against that State.
- The injured State must have called upon the State committing the wrongful act to discontinue its wrongful conduct or make reparations for it.
- The effects of countermeasures must be commensurate with the injury suffered, taking account of the right in question.
• Its purpose must be to induce the wrong doing State to comply with its obligations under international law, and the measure therefore must be reversible.

If a cyber-attack or cyber espionage violates the sovereignty of a State (which could include one of its private entities dependent on the State’s cyber infrastructure), and it has clear and convincing evidence that the wrong is attributable to a foreign sovereign State:

• The victim State may itself commit a wrong against the attacking State so long as the wrong is commensurate with the initial wrong (proportionality), and the response is aimed at inducing an end to the initial wrong (necessity) or provision of damages.

• Given that in most cases the evidence that a foreign State is behind an act of cyber wrongdoing will be found after the act is complete and damage is incurred, most countermeasures addressing cyber wrongs will be a demand for money damages.

• Thus a victim State should be able meet the elements of lawful countermeasures in the same way it would for a trade injury [2].

Countermeasures may only be taken by States, however States are entitled to outsource the taking of lawful cyber actions to private entities but when they do so, the States shoulder the legal responsibility for the actions [12].

C. Non-binding International Standards and Guidelines

After the 2007 cyber-attack on Estonia, NATO agreed to Estonia’s proposal to create a Cyber Center of Excellence. Between 2008 and 2012 the NATO Cooperative Cyber Defense Center of Excellence (CCDCOE) in Estonia produced the “Tallinn Manual on the International Law Applicable to Cyber Warfare”. The Tallinn Manual is an academic non-binding study on how international law applies to the most disruptive and destructive cyber operations that qualify as armed attacks (i.e. cyber warfare).

Every day, States are subjected to attempts to breach their network by malevolent cyber operations that do not rise to the level of an armed attack. Consequently, in 2013 CCDCOE started hosting a committee tasked with developing Tallinn Manual 2.0. This document expands analysis to international law during peacetime. Scheduled for completion during 2016, Tallinn Manual 2.0’s development process currently consists of drafters and editors creating a rough draft from analysis of treaties and practice for review by over fifty peer reviewers around the world. Unlike its predecessor, Tallinn Manual 2.0 is not academic in nature and is practice-oriented, citing treaties and how rules apply in practice [13].

D. Hack Back Examples

The following are two real world examples of what can happen when “hacking back” against the “attacker.”

1) Georgian Ministries and Banking Hacker: On March 2011 the Ministry of Justice of Georgia launched an investigation to find the perpetrators of continual persistent cyberattacks that stole confidential information from various government agencies, parliament, banks, and nongovernmental organizations (NGOs). The Georgian government’s Computer Emergency Response Team uncovered the cyberattacks that planted malicious software on a number of Georgian websites. The software was sophisticated and only installed on pages that “would interest targeted individuals”.

As the severity of the attacks increased over the course of 2011, Georgia allowed a computer to be infected on purpose, and placed in it a ZIP archive named “Georgian-NATO Agreement.” The file tricked a lurking hacker into downloading what he thought was sensitive information, and when opened the file installed the investigator’s own malicious code on the alleged hacker’s computer. With the code installed the investigators rapidly mined the suspected hacker’s computer for sensitive documents. In one Word document, the investigator found instructions on how to hack particular targets; as well as website registration data linked to an address within Russia. Additionally, the investigators were able to use the webcam on the Russian hacker computer, took his photo, and then published several images of him [14].

2) Blue Security’s Blue Frog Anti-Spam Service: During 2005, Blue Security offered consumers Blue Frog anti-spam service that went beyond filtering or blocking spam. When a user signed up for a Blue Frog account, the service provider directed the user to install a piece of software on the client’s personal computer and could list up to three email addresses on Blue Security’s Do-No-Intrude Registry. Blue Security in turn built a community-based anti-spam system which would try to persuade spammers to remove community members’ addresses from their mailing lists by automating the complaint process for each user as spam is received.

If community members reported spam to Blue Security, the report was analyzed to make sure it met their spam guidelines, then Blue Security would notify sites sending illegal spam to Internet Service Providers which hosted them, to other anti-spam groups, and law-enforcement authorities to get the spammer to cease and desist. For each spam message a user received, their Blue Frog client software would send the spammer one generic complaint, including instructions on how to remove Blue Security users from future spam campaigns. Blue Security operated on the assumption that as the community grew the flow of complaints from tens to hundreds of thousands of computers would apply enough pressure to convince spammers and their clients to stop spamming members of the Blue Security community [15].

In May of 2006 the Blue Security company was subject to a retaliatory Distributed Denial of Service attack by spammers that led to collateral damage on the Internet. Blue Security eventually closed down its anti-spamming operations [16].

E. Challenges in Hacking Back

As can be seen from the “Blue Frog Anti-Spam Service” example, even when you can attribute with certainty a spammer (or cyber attacker) the outcome of hacking back may not always be positive. In many cases private actors lack access to the sophisticated attribution tools and information available to the government, and risk hacking back against innocents [10].
Given that many cyber-attacks cross international borders, there is also the challenge when the cyber-attacks on U.S. companies are criminal activities originating from another nation whose law enforcement organizations cannot or choose not to investigate and/or to arrest the responsible criminals. Does this failure or inability to act invoke the Laws of Armed Conflict? To determine if this threshold is met we must transition our discussion to the Military Framework and the applicable Laws of Armed Conflict which would govern any response under this criteria.

IV. FRAMEWORK FOR THE MILITARY APPROACH

Now that we have completed our examination of cyber-attacks in the framework of criminal law, we can move on to the examination these actions under the framework of military action. Conflicts between societies date back thousands of years. The earliest conflicts took place between tribes seeking to assert dominance over their neighbors or for control of particular pieces of land. The rise of nations led to the expansion of war to a point where it encompassed the entire world twice in the last century. The growth of war led the nations of the world to establish the laws of war that define what justifies armed response [17], [18]. The codification of the rules of war are collectively referred to as the Law of Armed Conflict (LOAC).

With the passage of time and advancements in technology, militaries of the world expanded the warfighting domains beyond land to include the sea and air. The global expansion of the Internet, and the reliance on information systems as part of national infrastructures led to some national militaries recognizing cyber as the newest of the warfighting domains in the last decade [17], [19]. Applying the military framework and accepting that this is a warfighting domain, conflict in it must abide by LOAC.

We will begin with an outline of the primary legal sources for LOAC. We will conclude with an examination of the principles borne from these sources which are relevant to our present discussion. The primary sources for LOAC that we will examine include The Hague Conventions, The Geneva Conventions and Geneva Accords, and the United Nations Charter (UN Charter). The principles that we will examine are Jus ad Bellum, Jus in Bello, neutrality, and the rules applying to non-state actors and non-combatants.

At the conclusion of this section we will examine how the concepts of Rules of Engagement (RoE) and Escalation of Force (EoF) are employed in the physical domain and then extrapolate them for application in the cyber domain. Our goal to illustrate how the ethics of the Military Framework may be employed to ensure ethical actions even by those who might not be bound by the legal implications of LOAC.

A. Sources of LOAC

The primary sources for LOAC are found within Hague Law, Geneva Law, and the UN Charter. It is important to note that the application of LOAC no longer requires a formal declaration of war [20]. For example, LOAC applied during the Gulf War, in the Falklands, in Sri Lanka, and Chechnya [21]. While LOAC was once referred to as the “Rules of War,” nations no longer refer to them in this manner for two reasons. First, under the UN charter adopted after World War II, the declaration of war was prohibited [21]. Second, it is the deliberate intent of the Geneva Conventions’ authors to cover a complete range of situations and to avoid legal arguments over the exact definition of war [21].

1) The Hague Conventions: The Hague Conventions of 1899 and 1903 culminated in the participating members signing treaties in 1899 and 1907 [17], [18], [20]. The Hague Conventions contribute to LOAC in many ways but of significance in the current discussion are the rules for respecting state neutrality and those which focus on limiting suffering during war [20], [22]. Particularly, minimizing the suffering of non-combatants [20], [23].

2) The Geneva Conventions and Geneva Accords: The Geneva Conventions of 1949 and the Protocol Accords of 1977 revised the earlier Geneva Conventions of 1864, 1906, and 1929. The conventions of 1949 see the shift from the concept of laws of war to the laws of armed conflict [20]. As DiMeglio et al. observe, the change “emphasize[d] that the application of the law and prescriptions did not depend on either a formal declaration of war or recognition by the parties of a state of war [20, p. 8].”

3) The Charter of the United Nations (1945): Adopted following the conclusion of WWII, the charter establishes the international body, known as the United Nations, to aid in the peaceful resolution of issues rising between nations. Article 2 of the charter outlaws the use of force by one nation against another. However, Article 51 allows a Nation to resort to force in the act of self-defense, and the protection of its citizens [21].

B. The Principles of LOAC

From primary sources mentioned above we derive several key principles for LOAC. The intent of these laws is to ensure the integration of humanity into war, preserve the fundamental human rights of persons who fall into the hands of the enemy, and to assist in the restoring of peace [20]. These principles encompass many important areas we need to consider when evaluating cyber-attacks under the auspices of armed conflict and while considering the ethical appropriateness of hacking back. As outlined previously we will examine the principles of Jus ad Bellum, and Jus in Bello, Neutrality, and Non-combatants.

1) Jus ad Bellum: Jus ad Bellum defines the conditions under which nation states can resort to war or the use of armed force [18], [19], [23]. Prior to the mid-19th Century there was no meaningful concept for Jus ad Bellum as a nation’s right to resort to force was acceptable and unchallenged [23]. There are five criteria which must be met in order for Jus ad Bellum to be satisfied. They are legitimate authority, just cause, right intention, last resort, reasonable chance of success, and proportionality [18].

Legitimate authority is a function of a nation’s sovereignty [18], [24]. When we evaluate this in terms of a state’s cyber infrastructure, the state’s sovereignty is demonstrated by its ability to control the infrastructure through legal or regulatory means [19]. Just cause is the intent to restore peace which cannot be achieved through other means [18]. Reasonable chance of
success is simply the probability that the action will have the intended outcome. Proportionality is defined as what can reasonably be foreseen as an outcome of an event [18]. As Dinstein observes, excessive is determined based on the ‘good faith’ expectation not the actual outcome [25]. Conversely, if an attack could reasonably be expected to cause excessive damage to the power grid of an entire city, the fact that power was lost in only one section of the city would not legitimize the attack [25].

2) Jus in Bello: This legal concept pertains to the proper conduct of individuals involved in war [18], [20]. The origins of Jus in Bello can be traced beyond the primary sources mentioned above to the mid-19th Century [20], [23].

Jus in Bello incorporates three basic criteria: discrimination, military necessity, and civilian due care. Discrimination is the ability to clearly delineate between military or just and civilian or unjust targets [18]. Military necessity requires that the degree and kind of force used to achieve a military goal is legitimate for the purpose of the conflict at hand [26]. Finally, civilian due care refers to minimizing the harm that might befall civilian persons and property [18].

Applying these rules to our current consideration, it would be illegal and unethical for those defending information systems to target systems that are not directly involved in the persistent attack against the affected systems. As stated earlier regarding The Hague and Geneva conventions, combatants should always seek to do as little harm as possible to non-combatants or civilians. In his work on the ethics of cyber operations, Barrett suggests that while the Stuxnet virus passed through both civilian and military systems it only impacted military systems [18].

C. Neutrality

Neutrality is a long standing concept both in terms of international relations and international conflict. A neutral state is any state not a party to an international armed conflict [18], [19], [24], [27]. This concept of a third party state’s neutrality adapts seamlessly with operations in the cyber domain. However, one of the fundamental historic concepts of neutrality is that a state’s sovereignty is not infringed by moving resources through its territory [20]. The movement of cyber weapons through interconnected devices cannot guarantee that they will not pass through devices that belong to a neutral third-party [18], [19].

D. Non-Combatants

The traditional definition of a combatant required that he be easily identified by the wearing of a uniform and/or insignia [19], [28]. These uniforms made it easy to distinguish them from non-combatants. The means of distinguishing them from non-combatants was intended to minimize harm to non-combatants and their property [20]. However, as DiMeglio et al. point out, the Additional Protocol I to Geneva Law amends this requirement to “only requires combatants to carry their arms openly in the attack and to be commanded by a person responsible for the organization’s actions [20, p. 74].”

E. Rules of Engagement and Escalation of Force

In order to ensure that LOAC is followed by combatants in the physical domain, RoE are established by senior leaders to enforce a standard of conduct of their subordinate fighting forces. The goal of RoE is to provide warfighters with guiding principles about when and how they can engage the enemy [29], [30]. RoE are written rules designed to limit the use of unnecessary force, causing harm to innocent individuals, and avoiding an escalation of violence [29], [31].

In order to ensure that RoE are followed, leadership will often employ guidelines for the EoF [29]. EoF are steps outlined by leadership with the intent of proportional response to a situation. For example, when a unit performs a mounted patrol in a town or city, they can mount signs on their vehicles warning other vehicles to remain 50 feet back from the convoy or to risk being shot. Should a vehicle behind the convoy ignore the sign, the next step in the escalation of force might be the use of shouted warnings, and hand-and-arm signals or shining lights at the driver of the vehicle in warning. If these actions fail to dissuade the approaching driver, a weapon system could be pointed at the vehicle in conjunction with more verbal warnings. Only after these steps, with the possible addition of warning shots or other deterrents (e.g. shots to disable the vehicle), do the RoE and EoF rules allow a Soldier or Marine to lawfully respond with lethal force against the approaching driver.

F. Applying RoE and EoF to Cyber Defense

This same approach can be applied when dealing with hacking back in cyberspace. In this scenario passive defensive options would be the first step in the RoE or EoF. This could be done through warning banners or through some other communication that informs the offender to cease and desist less he be subject to a more severe response (including arrest). The next step in escalation might include the rejection of packets and the employment of forensic measures to identify the source of the attack. Because attackers will often employ third party computers and networks to execute their attack, the next step could involve contacting the owners of the affected systems and to assist them in remediating their systems.

If all these efforts fail, then and only then, should defenders consider more drastic counterhacking responses as ethically justified. By following this approach, the network defenders are ensuring that the priorities of proportionality and minimizing harm to innocent third parties are followed.

V. Conclusions

When we began our journey we assumed, for the purposes of our discussion, that attribution could be proven in our hypothetical cyber-attack. We understand that this is a significant leap to make even with the challenge that attribution has demonstrated in the sample-size of state-vs-state cyber-attacks that is definitively known. In the cases of Estonia in 2007 and Georgia in 2008 there is no proof that the Russian nation was involved in the digital attacks that took place in those countries [32]. While there are theories that the United States and/or Israel was behind Stuxnet it cannot be proven [33].

With attribution taken out of the equation the crux of the ethicality of hacking back rests on the tests of proportionality...
and potential for civilian harm. Understandably, there are many lessons that can be taken away from the Law of Armed Conflict which can inform decision makers about the responsible use of an active cyber defense.

ACKNOWLEDGMENT

The authors would like to thank Dr. Melissa Dark, Dr. Ida Ngmabecki, Dr. J. Eric Dietz, and Dr. Eugene Spafford of Purdue University for their support in our first publication.

REFERENCES