A Fine-Grained Approach for Analyzing Active Cyber Response by the Private Sector

© 2014 James Denaro, Irving Lachow, and Randy V. Sabett

Abstract: This paper proposes a framework for classifying current active cyber response (ACR) techniques and a fine-grained approach for analyzing active cyber response by the private sector according to a rules-based analysis of those techniques. There is increasing acceptance of the role of the private sector in protecting its networks without resort to governmental authority, especially with respect to investigational or forensic techniques that do not disrupt target systems. Currently, there is no clear framework for assessing whether and when the private sector should be permitted to use even those non-disruptive techniques, as well as more disruptive active cyber response techniques to protect and defend its networks.

A diverse set of ACR tactics, techniques, and procedures (TTPs) are now available to the private sector, ranging from minimally invasive forensic collection to significantly disruptive cyber operations. Current paradigms for analyzing ACR by the private sector typically consider most or all ACR techniques as an undifferentiated group of illegal activities. Those approaches ignore the sometimes significant variances among the diverse techniques that may be used in ACR by the private sector.

This paper proposes that ACR for the private sector be defined as a diverse set of TTPs that can be (a) used for identifying, detecting, analyzing, and mitigating threats to a network and (b) classified along a spectrum of varying risk and permissiveness. The paper begins by reviewing the current models for characterizing ACR and proposing a classification method for ACR TTPs. The paper further proposes a decisional framework for determining whether a proposed ACR operation should be permitted to proceed based on factors including: severity of the threat; nature of the attacker system being targeted by ACR; effects of the ACR on the attacker system; time elapsed between detection of attack and initiation of ACR; and possible collateral effects of the ACR. The paper proposes that such a fine-grained approach to authorizing ACR by the private sector could be used by a cyber operations court or other authority hearing applications for ACR and issuing determinations regarding specific actions.

Keywords: fine-grained framework, active cyber response, private sector

1. Introduction

Setting: a typical Friday in corporate America. Just after midnight, the Chief Information Security Officer (CISO) of the Acme Corporation receives a phone call. One of her system administrators (sysadmins) reports that a possible malicious hack has been detected. She hurriedly drives to the office, and meets the sysadmin. They find that an active intrusion is under way and begin executing their incident response plan and assessing the damage. Attackers seem to have penetrated many segments of their internal corporate network, potentially exposing personally identifiable information (PII) and the company's most sensitive intellectual property (IP).

After a brief triage process, they weigh their options. Based on what they have found, completely shutting down the network is not realistic. Law enforcement cannot help immediately because they are swamped and Acme has yet to conduct a full damage assessment. In order to improve its defenses and possibly take affirmative action to stop the attack, the company wants to act immediately to gather as much information as possible about the source and intent of the attack. It is uncertain, however, what actions are legal or illegal under applicable law and no clear principles exist to guide the decision-making process. Further, Acme's lawyers have heard that any "active" steps could be viewed by the authorities as violating certain laws (particularly the U.S. Computer Fraud and Abuse Act (CFAA)).

While our tale of Acme is fictional, the dilemmas it poses are real. For example, Acme could embed certain functionality in its electronic assets and track where those assets are, when they have been accessed, and (potentially) allow certain actions to be taken. The key challenge is not technology; rather, it is the uncertain legal and policy climate in which the actions would be taken.[1]

A robust debate about the permissibility of active cyber defenses continues to escalate. The topic receives attention from Congress and new companies are being created with a focus on providing ACR services. However, despite these developments, there is still a lack of consensus on what specific actions a company can take and under what circumstances. [2]

The current set of applicable laws does not provide a clear answer on what is legal and what is not. Congressional attempts to address the issue have often been highly controversial and the debate polarizing. Even modest Congressional proposals, such as provisions in the Cyber Intelligence Sharing and Protection Act ("CISPA")[3] limiting liability for relying on and taking action on threat data, have received strong negative reactions.[4] In this paper, we will attempt to

[1] Many in the private sector view such actions as a necessity. For example, one expert has observed that "[w]e can prevail only if we mount near-perfect defenses . . . This . . . is, quite simply, too hard. A wholly passive strategy almost never works in the real world." Stewart A. Baker, *Skating on Stilts: Why We Aren't Stopping Tomorrow's Terrorism,* at 228 (2010).

[2] Much has been written on ACR and related activities when carried out by the government or military. *See, e.g.,* "Commission On Offensive Information Warfare; National Research Council Of The National Academies; Technology, Policy, Law, And Ethics Regarding U.S. Acquisition And Use Of Cyberattack Capabilities" (William A. Owens *et al.* eds., 2009). This paper, in contrast, focuses solely on private sector use of ACR. Less has been written about private sector actions, which are the focus of our paper.

[3] *See* https://www.govtrack.us/congress/bills/113/hr624 for the current status of CISPA.

[4] *See* "Statement of Administration Policy" at http://www.whitehouse.gov/sites/default/files/omb/legislative/sap/113/saphr624r_20130416.pdf (stating that the President's senior advisors would recommend that he veto the bill). *See also* "CISPA Passes

put forth a mechanism for helping to consistently determine what could be advisable.

We begin by reviewing some existing frameworks applicable to ACR. We then propose a framework based on fine-grained distinctions between different tactics, techniques, and procedures (TTPs) to determine what may be advisable and what may not be advisable in any given situation under the current legal and policy landscape.

2. Current Mechanisms by Which ACR Techniques May Be Assessed

For this paper, we use the phrase active cyber response (ACR) to mean a diverse set of TTPs that can be: (a) used for identifying, detecting, analyzing, and mitigating threats to a network; and (b) classified along a spectrum of varying risk and permissiveness. We derived this from a Department of Defense (DoD) publication defining "active cyber defense" as the "synchronized, real-time capability to discover, detect, analyze, and mitigate threats and vulnerabilities."[5] Thus, while many commentators often use the phrase "countermeasures," we view ACR as consisting of a wide range of TTPs that can be used to accomplish the discovery, detection, analysis, and mitigation described above.

House; Obama Threatens Veto" at http://www.usnews.com/news/articles/2013/04/18/cispa-passes-house-obama-threatens-veto and "Who Really Opposes CISPA" at https://www.eff.org/deeplinks/2013/04/who-really-opposes-cispa.

[5] Department of Defense Strategy for Operating in Cyberspace, available at http://www.defense.gov/news/d20110714cyber.pdf (July 2011).

Many approaches to analyzing ACR rely on the notion of a violation of the conceptual boundary between an internal, privately owned and controlled computer or computer network and an external, publicly accessible computer or computer network. For example, many stakeholders seem to believe that anything that doesn't fall neatly into the category of purely passive tactics constitutes "offensive" activity that is therefore illegal. We believe that such a view overly constrains potential uses of ACR.

A. Prohibitive legal concepts

Certain frameworks focus mainly on the aspects of cyber behavior that are prohibited. Involving both civil and criminal aspects, these "anti-hacking" laws outlaw actions based on unauthorized access to a network or computer system. These prohibitive legal concepts could be applied, under the current legal regime, to defend the attacker from many of the ACR that could be employed by the victim.

1) Computer Fraud and Abuse Act ("CFAA")

The CFAA[6] provides a statutory framework that often leads to a confusing set of analyses when applied to ACR. While many commentators agree that the CFAA was intended to be applied mainly to hacking situations, its lack of clarity (particularly around the undefined concepts of "access without authorization" and "exceeding authorized access") could lead to it being applied to an

[6] 18 U.S.C. § 1030. *See* http://www.law.cornell.edu/uscode/text/18/1030.

entity conducting ACR. It is precisely this lack of clarity that this paper is intended to address.[7]

2) *International law*

Differences and incompatibilities between international laws compound the problems associated with analyzing ACR techniques. While certain efforts have been undertaken to address such differences as they arise in a military setting, no such unified effort has occurred related to civil law.

On the criminal side, some commentators assert that the Budapest Convention on Cybercrime[8] would likely render certain ACR techniques illegal. For example, Article 5 relates to "System interference" and requires that "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 serious hindering without right of the functioning of a computer system by inputting, transmitting, damaging, deleting, deteriorating, altering or suppressing computer data." Thus, the Budapest Convention appears to direct ratifying states to enact laws potentially criminalizing a broad range of ACR techniques.

[7] For a discussion of the challenges associated with applying CFAA to ACR, see Irving Lachow, *Active Cyber Defense: A Framework for Policymakers*, (Washington, DC: Center for a New American Security, 2012).

[8] *See* http://conventions.coe.int/Treaty/en/Treaties/Html/185.htm.

B. *Permissive legal concepts*

1) *Self defense*

The notion of "cyber self-defense" is frequently raised in the context of ACR, often as a justification. Under such an analysis, analogies are drawn between existing law of self-defense and defense of one's network. While self-defense in one's home clearly enjoys legal protection, no such clarity exists in cyberspace. Legal uncertainties exist as to whether ACR as a form of self-defense would be permitted, particularly when considered in the context of the CFAA.

2) *Trespass*

The notion of trespass in the context of computer networks could be applied to an entity conducting ACR against an attacker. This tort gained popularity in the late 1990s in a series of cases that arose from, for example, excessive email being sent to CompuServe subscribers[9] and pornographic promotions being sent to an ISP subscribers.[10] The tort expanded in the early 2000's to such activity as screen scraping.[11] Courts began to reverse the trend established by these early cases when a California court established that a plaintiff cannot succeed as a result of "electronic communication that neither damages the recipient computer system nor impairs its functioning."[12] Such a distinction can be useful in the analytic approach proposed by our framework.

[9] *CompuServe, Inc. v. Cyber Promotions, Inc*., 962 F. Supp. 1015 (S.D. Oh. 1997).

[10] *America Online, Inc. v. LCGM, Inc*., 46 F. Supp. 2d 444 (E.D. Va 1998).

[11] *eBay Inc. v. Bidder's Edge Inc*., 100 F. Supp. 2d 1058 (N.D. Ca. 2000).

[12] *Intel Corporation v. Hamidi,* 30 Cal. 4th 1342 (Cal. S. Ct. 2003).

3) *Self-help in DMCA and UCITA*

Another framework concept by which one can analyze ACR involves the various statutory approaches that provide for self-help. For example, the Safe Harbor provisions[13] of the Digital Millennium Copyright Act ("DMCA") allow an online service provider ("OSP") to avoid liability by taking down allegedly infringing content that a third party has posted. The OSP can avoid liability to the poster for taking down the alleged offending content and also to the original copyright owner. Similarly, the Uniform Computer Information Transactions Act (UCITA)[14] allows certain activities that have a "self-help" aspect to them. Although only passed in two states, UCITA authorizes the use by software vendors of self-help mechanisms, including "electronic exercise without court order of a licensor's rights in the event of cancellation of a license because of a the [sic] licensee's breach of contract"[15] (*e.g.,* disabling software if the licensee doesn't pay).

[13] Section 512 of the DMCA has been informally designated as the DMCA Safe Harbor.

[14] *See* "Uniform Computer Information Transactions Act" available at http://www.uniformlaws.org/shared/docs/computer_information_transactions/ucita_final_02.pdf.

[15] *Id.*, at Section 816 "Limitations on Electronic Self-Help."

4) *CISPA approach*

A recent draft of CISPA[16] carries the self-help concept to a new level. CISPA focuses on cyber threat information and how it can be used. The controversy involves a liability safe harbor when cyber threat information is used in good faith. Specifically, the bill provided "[n]o civil or criminal cause of action shall…be maintained in Federal or State court against [a stakeholder], acting in good faith — (i) for using cybersecurity systems to identify or obtain cyber threat information or for sharing such information in accordance with this section; or (ii) for decisions made for cybersecurity purposes and based on cyber threat information identified, obtained, or shared under this section." The most recent version of the bill passed in the United States House of Representatives in April 2013, but was not voted on by the Senate.

C. *Existing analytical frameworks*

1) *Schmitt framework*

The Schmitt framework was created to characterize distinctions between permissible and impermissible cyber operations within the law of armed conflict (see below). The Schmitt framework does this examining seven factors: severity, immediacy, directness, invasiveness, measurability, presumptive legitimacy, and responsibility. It is recognized that evaluating these factors is an imprecise and subjective endeavor and that the factors are useful while not necessarily determinative. Moreover, the factors were not

[16] H.R. 264, available at http://www.gpo.gov/fdsys/pkg/BILLS-113hr624rfs/pdf/BILLS-113hr624rfs.pdf.

intended to be exhaustive, although they are often treated as such. Professor Schmitt has characterized the framework as being more useful for post hoc forensic analysis of particular cyber attacks than for characterizing real-time operations.[17]

[17] Stuxnet, Schmitt Analysis, and the Cyber "Use of Force" Debate, NDU Press, available at http://www.ndu.edu/press/cyber-use-of-force.html.

2) *Law of armed conflict*

The law of armed conflict (LOAC) applies to nation states and focuses on acceptable justifications for engaging in war (*jus ad bellum*) and limits to acceptable wartime conduct (*jus in bello*). LOAC can be a useful analogy in the context of ACR because it applies an effects-based analysis to assess what would be permissible under a given set of circumstances. For example, the International Group of Experts in the Tallinn Manual agreed that any cyber operation that caused harm to individuals or damage to objects qualified as a use of force, while operations that merely cause inconvenience or irritation do not qualify as uses of force.[18] For the U.S. Government, the physical effects are key and result in a set of evaluative factors about which State Department Legal Adviser Harold Koh stated: "[i]n assessing whether an event constituted a use of force in or through cyberspace, we must evaluate factors: including the context of the event, the actor perpetrating the action (recognizing challenging issues of attribution in cyberspace), the target and location, effects and intent, among other possible issues."[19] These factors can be useful in our proposed analytical approach.

LOAC also requires that combatants adhere to the principles of necessity and proportionality Necessity, in this context, requires that the ACR is necessary to defeat the cyber attack or prevent an imminent attack. This requires that mere passive defensive

[18] Tallinn Manual on International Law Applicable to Cyber Warfare, Rule 11, at 48.

[19] M. Schmitt, "International Law in Cyberspace: The Koh Speech and Tallinn Manual Juxtaposed," Harvard International Law Journal, vol. 54 at 19.

measures (*e.g.*, dropping connections, disconnecting systems from the offending network) would not have been sufficient. Proportionality requires that the ACR be no more disruptive to the target than necessary to accomplish the objective of terminating the cyber attack. The Tallinn Manual also supports the notions of necessity and proportionality in cyber defense in Rule 9[20] and Rule 14.[21]

3) *Hot pursuit*

In the context of U.S. criminal law, hot pursuit refers the urgent pursuit of a criminal suspect by law enforcement officers which gives rise to an exigent circumstance allowing police to enter private property without a warrant and possibly arrest the suspect. While this would generally be a violation of the Fourth Amendment prohibition on unreasonable search and seizure, the United States Supreme Court has adopted this principle (*e.g.*, in *Warden v. Hayden*).[22] The notion of hot pursuit also has applications to international law.[23]

In the context of U.S. civil law, hot pursuit permits the victim of a theft to make an urgent pursuit to recover stolen property from the thief. The victim may use minimal force to recover the property, but is not entitled to cause serious injury. If the thief responds with force, the victim may use a reasonable amount of force to defend

[20] Tallinn Manual on International Law Applicable to Cyber Warfare, Rule 9, at 36.

[21] *Id.*, Rule 14, at 61.

[22] *Warden v. Hayden*, 387 U.S. 294 (1967)

[23] Lionel Beehner, "Can nations 'pursue' non-state actors across borders?" 6 Yale J. Int'l Aff. 110-112 (2011).

himself, but must end the confrontation as soon as possible. If the victim is not in immediate pursuit, then the hot pursuit justification no longer applies.

Several cyber experts, most notably, Steven Chabinski, have argued that the notion of hot pursuit applies in the cyber realm.[24]

D. Existing frameworks necessary, but not sufficient

The concepts described above, while helpful when developing an initial analysis of ACR, do not sufficiently address the wide range of ACR techniques available. A purely legal analysis, for example, may address a specific technological approach but may not take into account the various subtleties when that approach is compared/contrasted with other similar techniques. Furthermore, each given set of circumstances may require a case-by-case analysis. Mary Ann Davidson, CSO of Oracle, advocated before Congress in 2009 for an approach to cyberspace that would be similar to the Monroe Doctrine, stating that ”the Monroe Doctrine did not detail the same intervention or even specific intervention for each perceived act of aggression. [It] merely laid out ‘here is our turf; stay out or face the consequences’ language that allowed great flexibility in terms of potential responses...With proper guidance, various government agencies and the private sector can find their natural role in guarding our cyber infrastructures in a framework similar to how we currently protect our real-world interests.”[25]

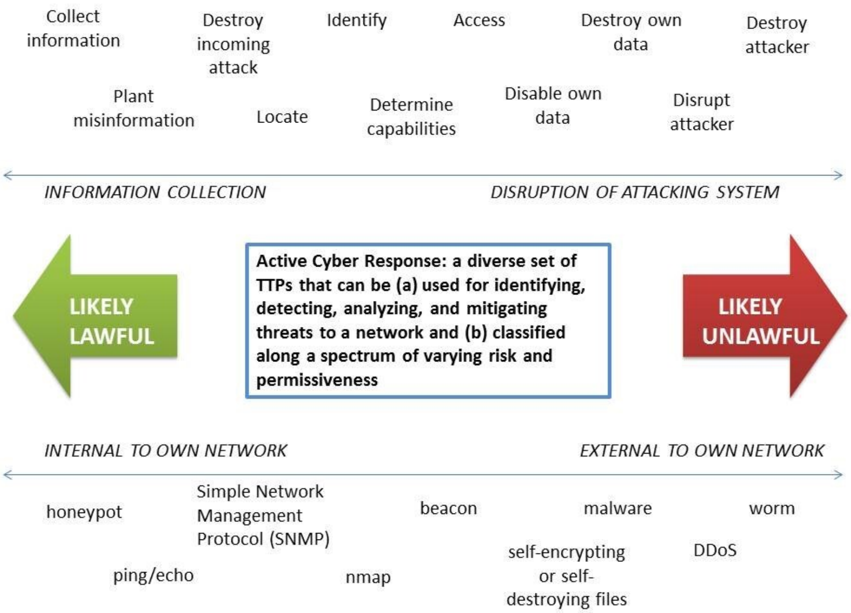
[24] For a brief reference to Mr. Chabinski’s arguments on this issue, see http://blog.cybersecuritylaw.us/2012/11/30/steven-chabinsky-crowdstrike-ex-fbi-cyber-division-talks-private-sector-cyberdeterrence-at-abas-natsec-law-conference/.

[25] Testimony of Mary Ann Davidson, Chief Security Officer, Oracle, March 10, 2009, to the Homeland Security Subcommittee on Emerging Threats, Cybersecurity and Science and Technology.

Ultimately, existing frameworks must be kept in mind when making near-term decisions about existing technologies because some of the principles are relevant in the context of ACR and certain laws are applicable. While such frameworks can inform risk-based decisions, ultimately such frameworks will likely prove inadequate. For example, the application of existing frameworks may result in conclusions that are inconsistent with each other. If an ACR is analyzed as a potential trespass on the attacker system, a victim of a cyber attack that is attempting to protect its intellectual property via techniques that "neither damage the [attacker's] computer system nor impair its functioning"[26] would not be found to commit the tort of trespass and would escape liability. Nonetheless, those same ACR activities may be held to be in violation of the CFAA if they included access without authorization or exceeding authorized access and criminal and civil penalties may be assessed on the victim. Thus, we believe existing frameworks alone are not helpful in analyzing various scenarios in a detailed way.

Consequently, private actors would benefit from a fine-grained framework that allows interested stakeholders to assess various actions on a case-by-case basis. To illustrate how an exemplary analytical outline might be structured, Fig. 1 presents a continuum of different techniques (along the bottom) plotted against general categories (along the top).

[26] *Intel Corporation v. Hamidi,* 30 Cal. 4th 1342 (Cal. S. Ct. 2003).


Collect information
Destroy incoming attack
Identify
Access
Destroy own data
Destroy attacker
Plant misinformation
Locate
Determine capabilities
Disable own data
Disrupt attacker
INFORMATION COLLECTION
DISRUPTION OF ATTACKING SYSTEM
LIKELY LAWFUL
Active Cyber Response: a diverse set of TTPs that can be (a) used for identifying, detecting, analyzing, and mitigating threats to a network and (b) classified along a spectrum of varying risk and permissiveness
LIKELY UNLAWFUL
INTERNAL TO OWN NETWORK
EXTERNAL TO OWN NETWORK
honeypot
Simple Network Management Protocol (SNMP)
beacon
malware
worm
ping/echo
nmap
self-encrypting or self-destroying files
DDoS

As can be seen from the sampling of activities and mechanisms shown in Fig. 1, a wide variety of ACR techniques exists. In order to use a fine-grained approach for analyzing ACR such as we are proposing, a mechanism for classifying such techniques will be needed. We look at such a classification process in the next section.

3. Classifying ACR Techniques

In order to assess the advisability of various ACR techniques, it is first necessary to understand the effects such techniques have, if any, on target systems. Without this grounding in the effects created by ACR options, it is not possible to accurately assess the legal and policy impacts such options can have.

ACR techniques can be grouped into two major categories: (a) those that collect information and (b) those that directly disrupt attacks. All of the techniques that we consider involve engagement with the adversary—that is what makes them active rather than passive—but there is still a broad spectrum of engagement. For example, one can create a honeypot to watch how adversaries behave when they think they are inside a victim's networks. One can watermark a document and then attempt to track its whereabouts after it has been exfiltrated. This approach can be taken one step farther by placing a beacon on a document that actively communicates its presence. One can also plant misinformation in the hopes that adversaries will change their tactics or make decisions in ways that benefit the victim.

The list below shows how a selected group of ACR techniques can be organized into the two categories we have just described:

- Information collection
    - Operations having effect solely within victim network
        - Honeypots
        - Watermarking
    - Operations having effect outside victim network
        - Beaconing files
        - Network commands (*e.g.*, ping, nmap)
- Disruption of attack

- Deny access to/from attacking system
  - Sinkholing
  - Blocking inbound and/or outbound traffic
- Disable or disrupt attacking system
  - Self-destructing files
  - Patch a vulnerability on a compromised third-party computer being used to launch attacks
  - Exploit a remote vulnerability on the attacker command and control node to stop the attack

The key factor to focus on when analyzing these techniques is the impact that they have on the attacker. The specific methods may change over time, but the types of effects that these actions may have on a system are consistent. For example, it is important to identify any ACR options that run executable code on an attacking system, independent of the nature of the code. The next section of this paper will provide a framework for assessing the permissibility and advisability of ACR techniques based on five key factors.

4. A Fine-grained approach for analyzing and authorizing ACR

In Section 2, we examined a number of existing frameworks that could be used to assess different ACR techniques, each with its own advantages and disadvantages. While the frameworks may be informative, they do not, by themselves, account for the technical details of active cyber response. In this section, we propose a fine-grained approach, informed by the strengths and weaknesses of the various frameworks discussed above, that can be used by lawyers, judges, technologists, policymakers, and others in the private sector to assess the risk of authorizing a given ACR technique. Our framework weighs five factors:

- Severity of the threat to the victim;
- Nature of the attacker system being targeted by ACR;
- Effects of the ACR on the attacker system;
- Time elapsed between detection of attack and initiation of ACR; and
- Possible collateral effects of the ACR.

Each of these factors is discussed below.

A. Severity of threat

The first factor in our framework is the severity of the threat to the victim system. It is fundamental that the magnitude of the threat facing a victim should be taken into account when determining what types of defensive actions are permissible. This is certainly true in the physical world, in matters both personal (*e.g.,* self-defense) and national (*e.g.*, law of armed conflict), and it should be true in the cyber world. For example, a "nuisance" attack that might embarrass a company, such as a website defacement, is very different from an attack that significantly affects both the technical and policy stances of a company by stealing (or threatening to steal) its most closely guarded IP. Logically, the victim in the latter case should be provided more latitude than the victim in the former case.

B. Nature of attacker system

The second factor in our framework is the nature of attacker system being targeted. This assessment can include several aspects, including the location of the system, the aggressiveness of the targeted system, its function, and its owner/operator. Location is important because taking actions that cross borders (be they state or national ones) may impact the legal regimes that apply and the government agencies that have jurisdiction over such actions. Function is important because certain types of systems, such as those that run industrial control systems, may be treated differently than others (*e.g.*, payroll systems). The level of aggressiveness can be used to help determine (or limit) the proportionality of the response. Finally, it may matter who owns/operates the system. For example, actions taken against systems within critical infrastructures may be handled differently than those that impact individual citizens. Public and private systems may also be viewed differently.

C. Effect of the ACR technique

The third factor that one must consider is the effect of a given ACR technique on the targeted system. The following questions may be relevant in conducting this assessment:

- Actions take place within one's own network
  - Do all actions take place on machines local to one's network?
  - To what extent are cloud resources used, if at all?
  - Does the ACR technique create any effects that impact systems outside of one's network? For example, is malicious code implanted on a document that is exfiltrated to an attacker's network whereupon it causes harm that network? Is any code used for ACR techniques executed on processors owned or operated by the network owner/operator (*e.g.*, the Internet Service Provider) supporting the victim organization?
- Actions require access to the attacker system being targeted
  - Is code being run on the targeted system?
  - Are any data being accessed?
  - Are any data being destroyed?
  - Are any systems being disrupted or destroyed?
  - Is availability being reduced?

Predicting the effect of a given ACR technique may be difficult because the same ACR technique may have different impacts on different systems. For example, a document that beacons out its location may have no impact on System A because that attacker is not monitoring his/her networks carefully. However, if the attacker running System B is taking precautions, a beacon on that system could cause actionable "harm" (such as disconnecting System B from the network) by running unapproved code that creates a policy violation.

D. *Time frame of response*

The "immediacy" of an ACR technique is an important factor to consider. In general, the more urgently that a response is needed, the more leeway that a victim may have in taking actions.[27] If it is known that law enforcement officials cannot or will not respond in a desired timeframe, companies may have more justification for taking actions on their own to gather information or perhaps stop an attack. This is analogous to the principle underlying "hot pursuit" in the physical world, which enables someone to run down and subdue a mugger even though such actions would be considered an assault under other circumstances. In contrast, if the threat facing an organization is minor or if law enforcement officials are able to respond to an incident quickly, then it may be harder to justify the need to take immediate ACR actions – at least some of them – on one's own.

[27] *See, generally*, "Adequate Attribution: A Framework for Developing a National Policy for Private Sector Use of Active Defense" available at http://digitalcommons.law.umaryland.edu/jbtl/vol8/iss1/3/.

E. *Possible collateral effects*

The fifth factor that must be considered when assessing an ACR technique is the likelihood that the action will affect innocent third parties. This aspect is similar to the “distinction” consideration that is central to the law of armed conflict. An ACR technique that is highly targeted on a specific attacker may be more acceptable than one that can inadvertently harm a number of innocent entities.

F. *Case studies*

The five factors described above can be combined in a weighted judgment to assess the advisability of a given proposed ACR action. We will now provide three examples to illustrate how our fine-grained approach can be used to assess ACR techniques. These examples will build on the case study that begins our paper. Case 1 will involve a fairly mild response with no collateral damage. Case 2 will be more aggressive and involve possible harm to a collateral system. Case 3 will involve a highly aggressive response that causes harm to the attacking system. We will describe each case and then provide a matrix that illustrates how our fine-grained methodology can be applied to these cases.

Case A

The Acme CISO needs to gather more information on what has occurred. She decides to establish a honeynet and populate it with documents containing false information. Her goal is to attract the cyber attacker to this honeypot in order watch its behavior and learn more about its motivations and the specific tactics, techniques, and procedures they use to launch its attacks.

Case B

Having observed the adversary in their honeynet, Acme has been able to determine how the attacker is entering their network. One of the Acme cyber analysts tells the CISO that as a result of running some SNMP queries, he knows of a vulnerability in the attacker's software that will let him access their command and control network. The Acme CISO tells him to go ahead and "hack" the software. The analyst is able to gain access the attacker's communications links and find the launch point of the cyber attack that hit Acme, which is a web server owned by a U.S. university. The analyst scans the contents of the web server's directory and is able to find several of Acme's most sensitive documents on that machine. The analyst thinks that the documents have not yet been copied from that server so if they are deleted, Acme's intellectual property may be protected. The CISO directs the analyst to delete the documents.

Case C

By accessing the attacker's C2 network and watching its behavior over a period of two weeks, Acme is able to determine that the cyber attacks hitting their network are originating from a competitor that is a state-owned company in another country. The Acme CISO meets with the company's senior leaders and they decide that they want to try and deter future cyber attacks from this competitor. The CISO directs her top cyber analysts to create an important-looking document that contains hidden code. If this document is opened on a system within the competitor's network, the code will execute and wipe the contents of that machine's hard drive. The analyst thinks that he can make this happen with 90% certainty, but cannot guarantee that the code will only execute on the competitor's systems; it is possible that the code could accidently execute on any of dozens of other machines comprising the C2 network, including a system owned by a "victim" (such as the aforementioned university). The CISO gives the OK to create the document and place it in the honeypot.

A summary of description of these cases aligned against the criteria in our framework is presented below:[28]

	Target System	ACR Effect	Time Frame	Collateral Damage
Case A	Internal	None	Immediate (hours-days)	None
Case B	University Web Server	System scanned and files deleted	Short-term (days)	Files wiped on "innocent" web server
Case C	Foreign State-Owned Company	Hard disk wiped	Medium term (weeks)	10% chance of HD wipe on dozens of systems

Analysis

Looking at Case A, it appears that the ACR actions taken by the Acme Corporation are both helpful and advisable. The company is gathering additional information in ways that do not affect others outside of the corporation. Its actions are justifiable and unlikely to create any legal or policy problems.

[28] In the summary table, the factor corresponding to the severity of the threat to the victim system is not presented because it is the same for each of the three cases.

Case B poses an interesting dilemma. On the one hand, it could be argued that the actions taken by Acme are not advisable. The company is accessing and destroying data that is sitting on a system owned by a third party, actions which are likely to be viewed as violations of the CFAA. On the other hand, one could argue that the actions taken Acme are justifiable. Acme's intellectual property was stolen and was found in a third party's system. If Acme believed that it had little time to notify law enforcement and wait for them to act before the IP was transferred from the third party to the attacker, then one could argue that this was a case of "hot pursuit." It is important that Acme is only destroying its own data on the university web server. If its actions cause collateral damage, such as destroying other files sitting on the web server, then the justification for this ACR choice becomes weaker. In the end, it is not clear if Acme's actions in this case are advisable. It would be necessary for Acme executives to weigh carefully the benefits and risks before proceeding.

There are numerous variants of this case that may be worth exploring. For example, what if Acme simply locates its stolen intellectual property and then notified the authorities but takes no further action? That case may be more advisable as it poses fewer legal risks, but it also increases the chances that the company will lose its IP to the attacker. Another interesting variant is if the data is being stored on the server of a critical infrastructure provider. In that scenario, the risks of collateral damage may also make it more advisable to simply find the data rather than attempt to delete it. The same may be true if the server storing the data is outside the United States, though the location of the server may make a difference, especially in terms of the laws of the nation in which the server is sitting.

The actions in Case C are not advisable. First, Acme is explicitly targeting a foreign-owned enterprise (FOE). This has several problems: it is illegal, it may have political repercussions for the United States, and it carries the risk of retaliation. Further, wiping an entire hard disk exceeds protecting its own stolen intellectual property. Acme's actions could be interpreted as escalatory and invite a strong FOE response. Given the timeframe in question, Acme could reasonably seek assistance in dealing with its adversary. Finally, Acme's actions pose a risk to the systems of neutral third parties. It is possible that Acme could accidently wipe the data from a computer in a critical infrastructure or other important facility (*e.g.*, a hospital). Overall, it is clear that Acme's actions in Case C are ill-advised and should not be approved.

5. Authorization of Proposed ACR Actions

One of the most significant criticisms of private sector ACR is that it would empower private entities to take unilateral action that could harm innocent third-parties. We propose that a new administrative body be formed for the specific purpose of applying the framework described above. The most appropriate mechanism by which to perform this analysis is a court of law.

This paper proposes that special court of limited jurisdiction – a Special Cyber Operations Court (SCOC)) – be created to hear these types of cases and rule on the permissibility of proposed ACR actions. The proposed SCOC process would require an application to the court seeking an order approving the proposed ACR action. The SCOC would then make reviewable determinations based on the framework described above. The SCOC could develop the technical expertise to apply the framework and act quickly, in real time, to approve or disapprove an ACR action proposed by an applicant.

The SCOC would need to have several optimizations. First, an entity engaging in ACR (or its representative) would need to take action on short notice and without exposing the planned action to the attacker. Thus, in addition to having sufficient technical competence to handle the issues raised above, the court would need to operate both (a) confidentially and (b) on an expedited basis. Because of its heightened expertise, this court is also better able to ensure that the interests of innocent third-parties are fully considered.

6. Conclusion

In this paper, we have proposed a decisional framework for determining whether an ACR operation should be permitted to proceed based on: severity of the threat; nature of the attacker system being targeted by ACR; effects of the ACR on the attacker system; time elapsed between detection of attack and initiation of ACR; and possible collateral effects of the ACR. Such a fine-grained approach to authorizing ACR by the private sector could be used by a cyber operations court hearing applications for ACR and issuing determinations regarding specific actions.

It is hoped that the creation of a Special Cyber Operations Court would resolve most of the concerns raised by the possibility of an ACR action. One of the primary concerns of those opposing ACR is that such actions could be undertaken by private entities having minimal supervision and few principles to guide their behaviors. Oversight by a technically competent and highly available judicial body charged with analyzing the ACR according to an established framework may solve some of these concerns.