ABSTRACT

The theft of intellectual property and classified data within the cyber domain poses a threat to the global economy and national security. In this paper, we discuss the concept of digital offensive countermeasures that the United States can use to defend its sensitive data and intellectual property, even after stolen data leaves U.S. Government networks. We analyze the plethora of legal and ethical issues involving the various degrees of invasiveness posed by such defenses against both foreign and domestic targets. The lack of established norms surrounding digital offensive countermeasures presents a unique duality in which such defenses may present a viable cyber deterrent for the United States but may also spark our next conflict.

INTRODUCTION

Intellectual property (IP) and sensitive data theft in the cyber domain poses a threat to the global economy and national security. For example, the $406.5 billion U.S. Department of Defense (DoD) F-35 Joint Strike Fighter program was the victim of multiple data breaches; moreover, the Chinese incorporated strikingly similar technology within their Shenyang J-31 stealth fighter which suggests that the United States (US) paid for the research and development for another country. [2]

The interconnectedness of businesses, governments, and the Internet makes it more appealing and viable for rival entities to reap immense technological boosts through IP theft. [3] Thus, corporations and governments must expand defensive efforts to make IP theft more difficult and costlier for attackers. Cybersecurity offensive countermeasures (OCMs) provide ways for achieving these goals. OCMs, also known as active defense strat-
egies, employ methods for achieving attack attribution, tracking intrusions back to their true source, and detecting attackers within networks.

Various OCMs are currently in use within the cybersecurity community, and they differentiate themselves along a spectrum of invasiveness. Honeypots represent the most benign end of the spectrum, in presenting attackers with a fake environment, permitting defenders to observe attackers’ tactics and techniques, and allowing defenders to create defensive signatures that can block future intrusion attempts. On the opposite end of the spectrum, we re-introduce the controversial concept of allowing adversaries to steal tampered IP that, when utilized, will result in physical destruction. In this paper, we discuss and explore the legal, ethical, and policy issues in a nation that protects its sensitive data with various OCMs.

1. RELATED WORK

This section outlines state-of-the-art defensive measures, provides a survey of threat modeling techniques, and introduces various offensive countermeasures that allow defenders to protect sensitive data and detect intrusions. These measures serve to increase the difficulty for successful cyber intrusions. As a general disclaimer, no single method can be effective against all threats, and practitioners should intertwine defensive measures when possible to reap synergistic effects.

1.1 State-of-the-art defenses

State-of-the-art defensive measures represent an evolution of defensive techniques within a cyber arms race: malicious actors exploit systems, defenders observe offensive methodologies through forensic analysis of attacks, defenders learn how to prevent such attacks from happening again, and attackers develop innovative ways to bypass defenses. We provide a cursory survey of current-generation solutions that
Lieutenant Colonel Jeffrey Biller is an Air Force Judge Advocate and Military Professor at the United States Naval War College in the Stockton Center for the Study of International Law, where he acts as the Associate Director for the Law of Air, Space, and Cyber Operations. The Stockton Center is the college’s research institute for the study of international law and military operations. Stockton Center faculty teach in the core curriculum and electives at the Naval War College, as well as in advanced international law courses around the world.

Lieutenant Colonel Jeffrey Biller is an Air Force Judge Advocate and Military Professor at the United States Naval War College in the Stockton Center for the Study of International Law, where he acts as the Associate Director for the Law of Air, Space, and Cyber Operations. The Stockton Center is the college’s research institute for the study of international law and military operations. Stockton Center faculty teach in the core curriculum and electives at the Naval War College, as well as in advanced international law courses around the world.

Lieutenant Colonel Jeffrey Biller is an Air Force Judge Advocate and Military Professor at the United States Naval War College in the Stockton Center for the Study of International Law, where he acts as the Associate Director for the Law of Air, Space, and Cyber Operations. The Stockton Center is the college’s research institute for the study of international law and military operations. Stockton Center faculty teach in the core curriculum and electives at the Naval War College, as well as in advanced international law courses around the world.

Lieutenant Colonel Jeffrey Biller is an Air Force Judge Advocate and Military Professor at the United States Naval War College in the Stockton Center for the Study of International Law, where he acts as the Associate Director for the Law of Air, Space, and Cyber Operations. The Stockton Center is the college’s research institute for the study of international law and military operations. Stockton Center faculty teach in the core curriculum and electives at the Naval War College, as well as in advanced international law courses around the world.

disrupt attackers’ methodologies, mitigate the effectiveness of their tools, and provide defenders with an improved security posture. The common theme throughout these defenses is that they all fail to assist defenders once an adversary has already attained access to intellectual property. Network administrators have implemented technologies commonly referred to as zero-client networks, where desktop workstations are replaced by virtualized systems to eliminate adversarial persistence but fail to protect stolen data. Zero-clients share a common secure baseline and exist only on demand; when a user logs in, a new version of the workstation is sent over the network. When the user logs out, the workstation is purged. This ephemeral characteristic requires attackers to continually re-infect targeted systems after each new session and prevents malicious actors from reliably using these systems within a botnet. If the secure baseline is frequently patched and assessed for vulnerabilities, zero-client networks can reduce the likelihood of external intrusions. This technology usually relies on network-wide databases for maintaining persistent information. These databases are not ephemeral nor is the data it stores; therefore, if an attacker can successfully exfiltrate sensitive data on these servers through an infected zero-client to an external destination, the zero-client technology is useless in protecting the stolen data. Thus, researchers have developed new technologies that aim to improve data access controls.

Zero trust networks (ZTN) represent an amalgamation of numerous permission-based security methods that can reduce external and internal threats from accessing sensitive data that zero-clients cannot protect. The basic concept behind ZTNs is that everything starts from an untrusted baseline and trust is established through a combination of methods, giving system administrators fine granularity control
over how devices and users access data. Using an up-to-date device inventory, ZTNs can deny or reduce access to resources if the requesting device is not using updated patches. ZTNs can restrict access based on time of the day or by where a user is logged in from. By layering permissions, ZTNs can prevent attacks from compromised administrator accounts, the most trusted accounts within a domain. Clearly, these security methods increase the difficulty for an attacker to access IP, but ZTNs cannot assist defenders if IP is exfiltrated to an external system. ZTNs can establish a chain of custody for accessed data within an environment through high levels of logging.

The Open Web Application Security Project (OWASP) maintains a series of best practices that guide network defenders toward implementing secure systems. Such guidelines provide an effective service towards building a more secure community and lends itself as input for various offensive countermeasures that we outline in the following section.

1.1 Offensive countermeasures

Offensive countermeasures and active defense strategies provide an interactive means for detecting and mitigating attacks. Honeypots are arguably the most benign OCM and present attackers with a fake virtualized environment. While attackers are attempting to exploit vulnerabilities and steal data, honeypots permit defenders to observe adversarial tactics and techniques and allows the creation of defensive signatures that can block future intrusion attempts. Honeypots serve as an immediate means for alerting defenders to intrusions because there is no legitimate use for honeypots only malicious actors will try to access them.

The concept of honeypots gave way to several innovations. Honeyports are fake open ports that detect network scanning and enumeration attempts by unauthorized personnel. Honeywords consist of fake passwords or password hashes that administrators seed within databases; if someone attempts to login using one of these honeywords, it triggers an alarm to defenders that the malicious actors have compromised the database and are attempting an intrusion. All of the methodologies of the honeypot family provide defenders with immediate notice and allow them to mitigate future incidents.

Web bugs are the first OCM we discuss that provides defenders with attribution for IP theft. Web bugs are beacons embedded within documents that alert a central server anytime those documents are accessed, allowing the central server to log the source location and time of access. If the malicious actors are not using a virtual private network or TOR, the central server logs their true location and defenders can begin to coordinate with law enforcement agencies for a formal investigation. Web bugs can be easily defeated if the malicious actor follows strict operational security and disables JavaScript within documents and uses location obfuscators at all times. Therefore, researchers developed new OCMs that would thwart attackers from discovering sensitive data.

Zip bombs are an OCM further along the invasive spectrum and are specifically designed
to conduct denial of service attacks on its victims and their storage resources. Zip bombs are crafted zip file archives that typically expand recursively and exponentially when an application unpacks it, or anti-virus application inspects it. The most famous zip bomb, 42.zip, is 42 kilobytes compressed and expands to 4.3 gigabytes; newer adaptations unpack infinitely until the victim crashes or the unpacking process is terminated. Coupling zip bombs with honeypots wastes the time of malicious actors while providing defenders with invaluable intrusion alerts.

These examples of OCMs present defenders with new methods for determining attack attribution and protecting intellectual property. When coupled with threat modeling techniques and other defensive strategies, defenders develop layers of security that increase the difficulty and cost of attacks for malicious actors. Nevertheless, these efforts may prove insufficient in the modern security environment and may require more overtly offensive methods. In the next section, we explore adaptations to traditional OCMs from this section that could lead to physical damage or destruction in the real world.

2. OFFENSIVE COUNTERMEASURES FOR NATION-STATES

In this section, we discuss various OCMs along a spectrum of invasiveness and discuss the possible legal implications of their use by state actors. The legal and policy concerns are driven both by the type of OCM to be employed and the context of their use. Therefore, we analyze each type of OCM first in their potential use against foreign targets, both within and outside of armed conflicts, and second, when the OCM has a domestic target. This article assumes that the state agency utilizing the honeypot has the appropriate foreign-intelligence, counter-intelligence, military, or law enforcement authority to conduct the operation utilizing the OCM.

We note at the outset that the application of law to the cyber domain is less than fully developed. International law often adapts slowly to new technology, as states create international obligations over time through a combination of formal agreements and customary law. There exist few formal international agreements related to cyber and customary law requires states to make formal pronouncements on their understanding of legal obligations. To date, states have been reticent to make such pronouncements on the application of international law to cyberspace, or even agree to basic international norms.

Several legal scholars have taken up the challenge and provided their interpretations of the application of existing law to cyberspace. The most in-depth and widely-cited of these efforts is the Tallinn Manual project, recently updated and expanded as the Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations. This manual primarily represents the opinion of nineteen international law experts, referred to in the manual as the International Group of Experts (IGE). Although this paper frequently references the Tallinn Manual 2.0 as an indication of scholarly opinion, we reiterate that only states can create international legal obligations.
2.1 Beaconing implants

The most uncontroversial OCM involves implanting sensitive documents with a “beacon.” This type of OCM is already in general use as a counter-intelligence technique. These types of beacons alert the owner anytime an unauthorized user accesses a protected piece of intellectual property or sensitive document from an unauthorized location. Beacons vary in implementation but consist of an embedded application or script that sends information to a centrally-controlled server. This server, in turn, logs metadata associated with the access request, such as the internet address of the host computer, its operating system, and time of access.

Depending on the beacon implementation, some embedded applications can facilitate additional intelligence collection on the miscreant’s system. This collection can involve data exfiltration from the system, keystroke logging, and access to physical devices such as webcams. This collection can be persistent and facilitate the installation of additional software.

Beacons are triggered any time a user opens an implanted document. This OCM is indiscriminate of the targets’ national origin or geolocation. Legitimate users also trigger the beacons and the centrally-controlled server logs metadata associated with authorized systems; US government acceptable use and monitoring policies authorize this type of data collection of employees.

2.1.1 Implications for foreign targets

The most accurate statement about the use of beaconing OCM against foreign targets is that their use is unregulated by international law. These OCM passively collect data from the affected system, pass it back to the original user and do nothing to affect the functionality of the target system. This passive collection is akin to espionage, which does not per se violate international law. The key factor is that this type of collection results in minimal degradation to the system. This analysis is consistent with Tallinn 2.0, where a majority consensus of the IGE believed that a beaconing honeypot, collecting data from foreign targets, does not violate international law during either peacetime and hostilities. Part of the IGE’s reasoning is that there is a sovereign right to protect sensitive data by embedding beaconing OCMs within sensitive documents stored within its borders. This sovereign right to protect data or code contained on a system within one’s borders is a key factor when analyzing OCM with more severe effects, as will be discussed in following sections. The IGE also found that any collection of data resulting from the beacon would constitute nothing more than cyber espionage. The collection of metadata is relatively benign, as it collects data that must be in unencrypted, plaintext form for proper transmission.

Examining the use of OCM for military purposes within the context of ongoing armed conflicts requires the application of International Humanitarian Law (IHL), which seeks to balance military requirements with the protection of the civilian populace. However, most
of the laws restricting military operations are unlikely to apply in the case of beaconing implants. Simple collection of data and subsequent computer network exploitation (CNE) does not inflict violence upon the enemy and therefore does not qualify as an “attack.” [30] Only actions that constitute an “attack,” which require an element of violence, are subject to the targeting provisions of IHL. [31]

Although the beacon itself does not constitute an attack, targeting cells might use the information gained from the beacon to identify military objectives for future attacks. During hostilities, malicious actors that trigger a beacon might identify themselves as a military objective if their activity triggering the OCM is determined to have a military purpose.32 Even if the malicious actor is not a uniformed member of the armed services, their military activities may, dependent on multiple factors, result in a loss of their immunity from attack under IHL. [33]

We next consider malicious foreign actors who trigger beaconing OCMs while performing criminal acts (e.g., stealing sensitive government data for commercial gain and not for reasons related to national security). Theft of information through cyber means violates several provisions of the US federal criminal code. 18 U.S.C. §1030 prohibits the intentional access of a computer without authorization to obtain information from the U.S. Government. Additionally, 18 U.S.C. §641 covers the theft of US property and information and does not discriminate based on the sensitivity of the stolen data. Criminal theft, however, does not necessarily equate to an internationally wrongful act for which a foreign state could be held liable. The bulk of international law applies to states. However, individual actions may be attributed to a state if the individual is acting as an agent of the state or under another theory of state responsibility. [34] Federal agencies will need to evaluate the evidence to determine whether to proceed as a counter-intelligence or law enforcement investigation, which may affect both procedural requirements and the permissibility of specific investigational tools. It is important to remember that several federal law enforcement agencies, including some belonging to the Department of Defense, may operate under both sets of authorities.

2.1.2 Implications for domestic targets

When beacons are triggered by unauthorized United States Persons (USP), domestic law, including the Fourth Amendment, the Electronic Communications Privacy Act (ECPA), and 18 U.S.C. §1030 may limit the uses of such beacons without appropriate court orders. [35] The Fourth Amendment protects individuals against unreasonable searches and seizures by the government and applies when a reasonable expectation of privacy exists. [36] Government systems typically provide a warning to those accessing the system that the access constitutes a waiver of the reasonable expectation of privacy.
ECPA lays further protections on privacy beyond the Fourth Amendment and requires specific types of warrants or court orders depending on the nature of the information to be accessed. Unlike the Fourth Amendment, which applies only to state actors, ECPA applies to private citizens as well. ECPA is the umbrella act for the Wiretap Act (18 U.S.C. §2511), Stored Communications Act (18 U.S.C. §2701), and Pen Register, Trap and Trace Act (18 U.S.C. §3121). Each of these acts protects different types of data in different ways, and any use of honeypots should be reviewed carefully to ensure compliance with these acts.

A significant exception common to all sections of ECPA is the service provider exception. These exceptions, such as 18 U.S.C. §2511(2)(a)(i), permit limited interception of data when service providers are engaged in activity “which is a necessary incident to the rendition of his service or to the protection of the rights or property of the provider of that service...” If the beacon is used solely to protect the service provider’s provision of services (to include the government when acting as a service provider), then such use of beacons will be permissible. However, the use of beacons to gather content data against the miscreant necessitates appropriate law enforcement or counter-intelligence authority to proceed. State laws, which vary widely in the area of privacy, may also impact the use of OCMs domestically.

Complications limiting the use of beacons often arise from unclear technical attribution. Given that hackers can obscure the source location of their operations using technologies such as secure tunneling or virtual private networks make it difficult to determine if the actor who triggered the beaconing OCM is a USP or not. If the collection is taking place within the US, then the agency should proceed under the assumption that the individual is a USP, until credible evidence reveals otherwise. \[37\]

A common criticism regarding the domestic use of honeypots by a law enforcement agency is entrapment. Entrapment is defined in Black’s Law Dictionary as “a law-enforcement officers’ or government agents’ inducement of a person to commit a crime, by means of fraud or undue persuasion, to later bring a criminal prosecution against that person”. \[38\] However, the use of OCMs by themselves constitutes neither fraud nor persuasion. Rather, they are relying on the miscreant’s initiative to access a protected system illegally. Should the law enforcement agency undertake additional acts designed to prompt the miscreant into accessing the system, then entrapment may be a factor.

### 2.2 Inert taint within honeypots

Next, we discuss the implications of the US intentionally tainting sensitive documents and plans for physical systems such that they become inoperable or inert when built. This type of OCM includes placing a tainted copy of a sensitive plan within an organization’s network, which consists of nuanced changes to the original schematic. These changes should be hard to distinguish by anyone outside the program developer. Tainted copies and original copies should exist on segmented systems so that a miscreant cannot easily exfiltrate both versions.
A theoretical example of this OCM is a government organization hosting the schematics for next-generation stealth aircraft technology that includes a flaw which makes the objects detectable to the originating state when integrated. Another example includes the theft of munition designs that do not fire or do not detonate correctly due to an altered wiring diagram. This OCM causes a miscreant to waste money during production, tarnishes the reputation of the intelligence collector, and causes adversarial organizations to second-guess the integrity of other stolen IP.

The government actor could also couple tainted sensitive data with a beacon to improve situational awareness of offending actors. Such a combination provides a state with the option to conduct follow-on CNE against its adversary and to monitor the subsequent chain of custody for stolen sensitive data. Chain of custody is key to an intelligence agencies ability to track and observe associated actors that receive the tainted data. If not incorporated, the affected governmental entity may lose visibility on the data once it leaves the network (which subjects this OCM to many of the same limitations of other defensive techniques we identified in Section 2).

**2.2.1 Implications for foreign targets**

The possibility of using an OCM with foreseeably harmful effects against foreign actors raises the difficult issues of legal attribution. As used in this section, attribution refers to circumstances when a state can be held responsible under international law for the breach of an international obligation by an individual. Actions which both breach an international legal obligation (i.e., the prohibition on the use of force found in the UN Charter), and are legally attributable to a state, are known as “internationally wrongful acts.” Typically, attribution refers to situations where the actions of an individual or non-state group become the responsibility of the state. Although this issue may come up regarding OCMs, the analysis will change little from other types of cyber operations. Actions can be attributed when the effects are determined to meet the standards of causation, required to hold a state responsible. Determining this type of attribution of the effects resulting from honeypot operations is particularly difficult due to the lack of clarity of the legal standards of causation.

Under general principles of law, effects are only attributed to a cause if they meet certain specific standards of causation. Potential standards include intent, foreseeability, strict liability, and proximate causation. Unfortunately, there is no agreement under international law, either in treaty or customary law, as to which standard would apply in the context of OCMs. Whereas the harmful effects may be foreseeable, and possibly even intentional, the honeypot was not the proximate (or nearest) cause of the damage because an intervening event, the data’s theft by the miscreant, had to occur for the effects to result. The *Tallinn Manual 2.0*’s discussion of honeypots reflects this uncertainty in the law. The IGE’s opinion was divided, with the majority holding that no attribution would exist for the state employing the honeypot, as the affected state had to take the affirmative, albeit unintentional, the step of transmitting the
tainted files into their own system. It must be stressed, however, that this is not an issue upon which states have yet to officially comment. Until the customary law develops in this area into codified, binding law, the legal question must be considered unresolved.

Legal attribution is particularly important when discussing OCM, due to the potential of affecting unintended targets. A potential weakness of the use of honeypots is that unforeseen actors may acquire the tainted material and use them in a manner which contradicts the purposes of the employing state. It may be wise to utilize protective devices, such as self-destruct mechanisms or the ability to remotely delete the file. However, given that the result is unlikely to be legally attributed back to the originating state for the above-stated reasons, this is more a policy than a legal concern.

Another factor determining the legality of OCMs is whether a state employs the honeypot for use in peacetime or within an armed conflict. During hostilities, if a state uses a honeypot for a military purpose, then states must examine the applicability of IHL rules. One such IHL rule that applies to all military operations within an armed conflict, whether or not that operation meets the definition of an attack, is the requirement that “constant care” be taken to spare civilians and civilian objects. Although the exact meaning of constant care is difficult to pin down, the Tallinn IGE states that the duty requires “commanders and all others involved in the operations to be continuously sensitive to the effects of their activities on the civilian population and civilian objects, and to seek to avoid any unnecessary effects thereon.” Therefore, if it is foreseeable that a triggered OCM will affect the civilian population with no military advantage to be gained, then commanders should seek to avoid or limit these effects if possible. The general principles of humanity and military necessity support the constant care requirement as well. As stated in the DoD Law of War Manual: “A cyber operation that would not constitute an attack, but would nonetheless seize or destroy enemy property, would have to be imperatively demanded by the necessities of war”.

An additional IHL obligation that potentially exists even where there is no legal attribution for an attack is the requirement, whether by treaty or by policy, to conduct legal reviews of weapons, means, and methods of warfare. A legal review examines IHL principles such as superfluous injury, discrimination, and explicitly banned arms to determine their potential compliance under IHL. More specifically under API Article 36 is the “obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party”. However, there is lack of agreement as to whether, or to what extent, the API requirement is considered customary international law. Although the US is not a party to API and has made no statement as to the customary nature of Article 36, by policy the US conducts legal reviews of weapons and, in some cases, cyber capabilities. The U.S. Air Force, for example, requires by policy legal reviews of “cyber capabilities,” defined as “any device or software payload intended to disrupt, deny, degrade, negate, impair or destroy adversarial computer...
systems, data, activities or capabilities”. In addition to an acquisition level legal review, all cyber operations that intend to produce effects that amount to an attack under IHL should be reviewed for compliance with targeting restrictions under IHL. For parties to the treaty, these requirements are encapsulated in API. However, many of the API rules are understood to constitute customary international law. These requirements include the rules governing distinction and proportionality.

There is much debate over what cyber effects qualify as an attack. The Tallinn Manual 2.0 definition of a cyberattack is “a cyber operation, whether offensive or defensive, that is reasonably expected to cause injury or death to persons or damage or destruction to objects.” The rule is facially uncontroversial. However, what constitutes “damage or destruction” to objects is complicated by operations that cause a system to cease functioning or reduce functionality without any apparent physical damage. Whether such effects against functionality result in qualification as “damage” has yet to reach a consensus under international law.

The use of OCM whose effects would normally qualify as an attack in an armed conflict raises a bit of a paradox. Despite potential legal or policy requirements for acquisition-level legal reviews on cyber capabilities, the use of such a capability in an OCM may not require consideration of IHL targeting provisions. As previously discussed, the effects of an OCM may not be legally attributed to the originating state because it is not responsible for the transmission of the code outside its system. Furthermore, it is unlikely to be legally attributed to the state that triggered the OCM because that state would be unaware of the OCM. This lack of responsibility creates the unsettling situation where a near-total lack of responsibility exists. The only definitive legal restriction on use continues to be the requirement to take “constant care” to spare the civilian population, civilians, and civilian objects. Tallinn 2.0 requires commanders “to be continually sensitive to the effects of their activities on the civilian population and civilian objects, and to seek to avoid any unnecessary effects thereon”. As the only relevant use restriction, commanders employing these OCM should be aware that any impact on civilians is likely to be held as their moral, even if not legal, responsibility.

2.2.2 Implications for domestic targets

The potential for OCM to affect unintended domestic targets with kinetic effects, such as those resulting from an inert taint, implicates a complex array of laws affecting the domestic use of information operations. In addition to intelligence oversight laws restricting collection against U.S. persons (USPs), several federal laws and regulations limit the domestic use of information operations. An agency not operating under law enforcement, or counterintelligence authorities must be prepared to react if they receive an indication that their OCM operation has affected a USP. Depending upon the situation, this may involve cleansing procedures, termination of the operation, or handing over to an agency with the appropriate authorities. However, this would not limit criminal responsibility for the unauthorized user which accessed tainted documents on restricted government systems. If a USP is attempting
to retrieve data from a system without authorization, 18 U.S.C. §1030 is again relevant for prosecution, as well as laws restricting the gathering of national security-related information (e.g., 18 U.S.C. §793). Should that data then be traced on to an additional user, such as a foreign government, illegal disclosure laws such as 18 U.S.C. §798 may also be relevant. Furthermore, government actors, who may have authorized access to the system, are barred from illegal removal of classified material under 18 U.S.C. §1924. Two additional considerations for domestic targets are the potential for liability if the tainted documents result in damages and political blowback if the use of the tainted documents results in a threat to public safety.

2.3 Tainted honeypots for subversion

Unlike the inert taint we previously described, this OCM involves tainting sensitive data so that a cyber capability on the adverse system can become controlled by US forces when activated. Expanding upon a previous example, this OCM is possible if a government entity taints application source code embedded within a plan related to stealth technology. This new code base would allow US entities to actively change target systems, such as disrupting or outright commandeering an aircraft that utilizes the stolen intellectual property which contained the embedded code.

2.3.1 Implications for foreign targets

There is a qualitative leap with this type of OCM versus the previously described variants. Whereas the previous OCM implanted either a passive beacon or tainted documents with no further active involvement from the creating state, this type of OCM allows for active involvement in the affected system. The ability to take actions after the OCM delivers the tainted code alters the calculus under both domestic and international law. Typically, computer network exploitation against foreign computer systems by US government entities is governed by signals intelligence authorities. However, this OCM potentially involves interactive manipulation of systems for non-intelligence collection purposes, which may necessitate offensive cyberspace (or defensive cyberspace operations - response action) authorities. Conducting operations under different authorities may alter governmental oversight responsibilities, funding limitations, and approval requirements. If a foreign person or entity is the OCM target for domestic law enforcement purposes, then foreign law enforcement cooperation is typically required, most often through Department of Justice procedures. \[56\]

When utilizing this type of OCM against foreign targets, there is no longer the legal attribution limitation discussed regarding passive OCM operations. If the OCM permits active involvement on the target system, such as the commandeering of a system, then the originating state becomes legally responsible for the effects that result once the state takes an active involvement in the target system. If the effects amount to an internationally wrongful act, such as interference in inherently governmental functions of another state or illegal use
of force, then the affected state could respond with actions in self-defense, take countermeasures, or demand reparations, depending on the nature of those effects.\[^{57}\]

### 2.3.2 Implications for domestic targets

The use of honeypots tainted for subversion against domestic targets is governed mainly by the previously mentioned electronic privacy laws, such as the Fourth Amendment and ECPA. However, if this OCM were to also deprive a USP of their rights to life, liberty, or property, procedural due process rights will also apply. These rules are discussed in greater detail below in the Lethal Honeypots section. One legal limitation that does not apply to government agencies engaged in official government functions are the prohibitions against unauthorized access contained in 18 U.S.C. §1030. These prohibitions have exceptions for “any lawfully authorized investigative, protective, or intelligence activity of a law enforcement agency of the United States, a State, or a political subdivision of a State, or of an intelligence agency of the United States.”\[^{58}\] Depending on the investigative techniques to be employed on the target system, the “lawfully authorized” investigation will likely require a warrant or other appropriate court order to employ the honeypot and utilize the access provided by the embedded code. Although the specific requirements for a warrant and its various exceptions are outside the scope of this article, it is recommended that the language contained in the warrant application be very specific about the intended actions during the exploitation.

We here note that there is a potential legal precedent for a US non-state actor gaining remote access and control over systems belonging to US citizens without their consent. In 2012, the Microsoft Corporation leveraged the Racketeer Influenced and Corrupt Organizations Act\[^{59}\] as the legal basis for their takedown of the Zeus family of malware.\[^{60}\] In this operation, Microsoft gained remote control over systems infected with the Zeus malware and cleaned the infection from the systems. Microsoft claimed that the Zeus malware posed an imminent threat to the general public. It was key that a federal court blessed the operation, undertaken in cooperation with federal law enforcement.

### 2.4 Poisonous honeypots

In this section, we discuss the concept of intentionally developing “poisonous” honeypots. These are OCM containing embedded code with the potential to levy destructive effects, including physical destruction or lethal effects. The OCM is activated when the target seeks out, steals, and utilizes (or consumes) tainted code, data or schematics. Poisoned systems are distinct from systems infected with computer viruses, which allow malicious code to transfer to other systems when it meets various conditions through a self-replicating mechanism. In poisoned systems, the target is responsible for acquiring and ingesting the tainted information/code and then acts as the replication mechanism. This OCM becomes potentially more lethal depending upon the actions of the target system operator.
Poisonous honeypots have proven effective as an OCM. In 2004, the U.S. Government declassified a covert Central Intelligence Agency (CIA) operation involving a “poisoned” Siberian gas pipeline. The CIA allowed Soviet spies to steal tainted pipeline control software, which when installed within their pipeline control systems caused an explosion that resulted in millions of dollars in damages.\[^{61}\] The explosion occurred in a remote location of Siberia and did not harm any humans. However, it is not a stretch to imagine poisonous honeypots that could potentially result in injury or loss of life. Fast-forward to 2007, the U.S. Department of Energy conducted a proof-of-concept cyber operation against a network-connected power generator that resulted in a controlled explosion.\[^{62}\]

Poisonous honeypots in this implementation are similar to a more conventional, but controversial weapon: booby traps. The Mines Protocol and Amended Mines Protocol define booby traps as “any device or material which is designed, constructed or adapted to kill or injure, and which functions unexpectedly when a person disturbs or approaches an apparently harmless object or performs an apparently safe act”.\[^{63}\] As pointed out in Tallinn 2.0, it is by no means certain whether and how booby traps might apply in the cyber context.\[^{64}\] Questions as to the applicability of the booby trap provisions include threshold questions such as whether code or data could constitute a “device.” Even should a poisonous honeypot be considered a booby trap, its use as such would only be prohibited in certain circumstances, such as when used with objects associated with medical or religious functions.

Should a state develop a poisonous honeypot, then it may have to pass legal review as described in the previous sections. Engineers or software developers should work together with legal experts to ensure this type of OCM can discriminate based on characteristics such as geolocation or biometric traits, for example, keystrokes.\[^{65}\] Tailoring this OCM to affect only a predetermined, legitimate target or groups of individuals also makes intentionally lethal honeypots more palatable and viable to government policymakers. A kinetic analogy for such a tailored OCM would be the use of landmines along the Korean Demilitarized Zone (DMZ).\[^{66}\] These mines are deployed in a defined and publicized area and are only intended to harm vehicles or personnel that violate known access restrictions within the DMZ.

Such an analogy leads to the following question: should and how can we effectively alert users that networks may contain intentionally lethal honeypots without the OCM losing its effectiveness? The Korean DMZ is delineated on a map and has numerous warning signs in its vicinity. It is unclear whether including honeypot warnings within electronic access consent banners would be an accurate translation within the digital realm. Additionally, the notion of “alert fatigue” renders many warning banners ineffective.\[^{67}\] Alert fatigue occurs when computer users are so inundated with innocuous warnings that serious warnings are bypassed and unobserved. Furthermore, it is unlikely that a computer hacker would heed a consent banner considering their objective completely violates any acceptable terms of use. However,
alerting a malicious actor to the presence of OCMs may achieve deterrence, or it may lead
the actor to place additional scrutiny towards any stolen data. We explore the latter in more
detail in Section 3.5.

2.4.1 Implications for foreign targets

The previously discussed analysis regarding state responsibility for the use of honeypots
holds for potentially lethal versions as well. However, it should not be dismissed that there
is a qualitative difference in the use of potentially lethal honeypots. Should the poisonous
honeypot work as intended, with lethal results on a foreign target during peacetime, states
may claim this result violates the Article 2(4) prohibition on the use of force. The Tallinn 2.0
rule defining the use of force in cyberspace is relatively uncontroversial: “[a] cyber operation
constitutes a use of force when its scale and effects are comparable to non-cyber operations
rising to the level of a use of force”. This definition includes “[a]cts that injure or kill per-
sons or physically damage or destroy objects...”. Although we have a situation where the
effects would normally constitute a use of force, we are led back to the same state responsi-
bility issue of attribution.

States should consider viewing lethal honeypot variants in the context of the entire UN
charter, particularly Articles 39, 51, and 53. There is a colorable argument that the use of a
lethal or physically destructive honeypot violates the overall purpose and intent of the UN
Charter. Even if no Article 2(4) violation is technically found, under Article 39 the “Security
Council shall determine the existence of any threat to the peace, breach of the peace, or act
of aggression and shall make recommendations, or decide what measures shall be taken....”.

The use of a poisonous honeypot, particularly one that appears out of proportion to the
information or system to be protected, may be considered by the Security Council to consti-
tute a “breach of the peace.” The Tallinn 2.0 IGE recognized this extended obligation in their
discussion of Article 2(4), postulating that “even acts that are not directed against either the
territorial integrity or political independence of a state may violate the prohibition when
inconsistent with the purposes of the United Nations”. While the UN Charter does prohibit the use of force, it does permit an exception when act-
ing in self-defense. Articulated in Article 51, it provides that “nothing in the present Charter
shall impair the inherent right of individual or collective self-defense if an armed attack
occurs....” This right is read by many states to include “anticipatory self-defense.” Although
definitions of this right abound, they all include some element of immediacy. If a state em-

ploys a lethal or physically destructive OCM to protect a very narrow class of systems, such
as those controlling vital elements of national security (e.g., nuclear command and control or
air defense), a strong argument exists that they are acting as a self-defense mechanism. The
immediacy requirement of anticipatory self- defense is met because accessing and tamper-
ing with such systems is strong evidence of another state’s intent to launch an armed attack
against the victim state, thereby permitting the use of force in self-defense. The use of honey-pots to protect vital national security systems may also be permitted as a “plea of necessity.” This customary law permits a state to take actions that would normally violate international law to respond to acts presenting a grave and imminent peril. This theory would require the defended system to be an “essential interest,” and the state must narrowly tailor the OCM to only protect “grave” threats against the system.

The argument for using lethal OCMs for anticipatory self-defense does not extend to defending a nation’s critical infrastructure and key resources (CIKR). CIKR is a domestic term that states utilize to define elements essential to security, public health, economy, and its overall way of life. CIKR does not hold any strict relevance to international law, and therefore a nation cannot integrate poisonous honeypots within CIKR defenses unless the CIKR is also a target implicating one of the previously defined theories.

When viewed within an armed conflict, OCM that may potentially inflict lethal or physically destructive effects on civilians must meet the “constant care” IHL obligation. Poisonous honeypots in an indiscriminate configuration have an increased likelihood of affecting non-combatants. As previously mentioned, poisonous honeypots can be manufactured to discriminate, selectively deploying against a target based on characteristics such as geolocation or biometrics. This obligation should not be read as being overly restrictive. Even the most carefully crafted OCM has the potential to affect a non-intended civilian target, and “constant care” is not defined in IHL. Instead, it connotes a general obligation of sensitivity to the civilian populace.

### 2.4.2 Implications for domestic targets

The use of poisonous honeypots domestically is highly restricted. The Due Process Clause of the Fourteenth Amendment prohibits the government from taking “life, liberty, or property without due process of law.” With regards to OCM designed to cause bodily injury or death, there are few scenarios that would not violate due process protections. Likewise, restrictions on liberty, such as the ability to communicate using a networked system, requires an order or adjudication from a court. However, it is possible that poisonous honeypots designed to destroy network equipment could be employed to protect certain narrowly defined systems, such as the previously mentioned national security or CIKR examples. The due process evaluation balances individual rights with government interests and may allow taking property without due process in limited circumstances.

The U.S. Supreme Court decision Mathews v. Eldridge established the factors for this balancing, holding that three factors stand out: First, the right to be impinged upon by the government action; second, the risk of depriving a right in error; and third, the burden additional procedural steps would take against the government interest. Given that a honeypot normally does not allow for procedural legal steps to be taken prior to affecting the property
rights of the miscreant operator, there needs to be an overwhelming government interest that would be unduly burdened by procedural requirements. Thus the limitation to the most critical network systems. Additionally, the effects of the OCM should be limited to system damage preventing the user from further accessing or harming the critical system.

2.5 Against using OCMs

Arguments cautioning against the use of OCM include both reasons of effectiveness and potential violations of the law. On the practical side, the value of hosting tainted schematics or code within a honeypot is diminished as valuable intellectual property existing alongside the tainted data may still be extracted by the unauthorized party. The tainted portions of the intellectual property should be nearly indistinguishable from the legitimate sections, but a highly-skilled technician may detect the taint before being affected by it. Early detection provides the technician with an opportunity to patch the stolen intellectual property in such a way as to restore original functionality. Thus, the mere creation of a tainted honeypot increases the likelihood of intellectual property theft.

Using the stealth aircraft example, a malicious actor that detects tainted elements of the data could integrate the stolen technology into their aerial platforms after conducting an abbreviated development period to repair data poisoned in the OCM. Similar to the theft of the F-35 plans, the actor now has a multi-billion-dollar capability at a fraction of the US research and development costs. The U.S. Government can take additional steps to mitigate this risk such as tainting a higher percentage of the IP or embedding more active forms of malware within the document. This controversial action could cause foreseeable harm to civilians.

Furthermore, OCMs, particularly lethal variants, may unnecessarily antagonize other states to such an extent that kinetic hostilities erupt. As we discussed in Section 3.4, OCMs that cause physical damage or induce casualties may be considered a breach of the peace, if not an illegal use of force. For perspective, consider a scenario in which a state successfully exfiltrates next-generation engine technology and integrates it within a “sixth-generation” airframe. During a test run of the newly-acquired technology, the aircraft crashes into an urban area because the foreign nation embedded a poisonous OCM within the data. An individual associated with the project leaks to the press that a foreign nation was the cause of the crash. How would the state react? This scenario breaks from the steady-state game of “spy-versus-spy” in which nations regularly conduct CNE and other forms of espionage against one another.

Also arguing against the prolific use of honeypots with effects ranging beyond espionage is the immature development, particularly regarding pronouncements by states, of international and domestic law as applied to the cyber domain. For OCM under international law, whether states can be held responsible for their effects is the threshold question. Currently,
the weight of opinion is against holding states responsible for employing OCM because the act of accessing a protected system and removing the tainted code, data, or plans is carried out by the miscreant. This is particularly true of honeypots that act in a more passive manner, such as beaconing. Such acts are unlikely to rise above the level of espionage, which is not per se regulated by international law. However, if the law develops a causation standard such as intent or foreseeability, those OCM employing more potentially violent effects could, at worst, be viewed as illegal uses of force violating the UN Charter, or, at a minimum, as breaches of international norms resulting in damaged international relations. Until international law matures in this area, developers of OCM should be careful to design them to be highly discriminate and with the minimum effects required to achieve their desired ends. These steps will also aid in ensuring their use within an armed conflict complies with IHIL.

Use of OCM domestically is more restricted than is the case against foreign targets and should also be given careful legal consideration. Even when employed by a government agency with the appropriate authorities, multiple areas of law restrict their use against USPs. Criminal law, privacy law, national security law, and due process considerations all limit when and how OCM can be employed domestically. Furthermore, public policy considerations such as public safety may limit the use of OCM, particularly those which may result in physical damage to objects or injury and death to persons.

3. CONCLUSION

Throughout this paper, we identify various OCM that state actors may use to complement threat modeling and other state-of-the-art defensive techniques. OCMs provide defenders with a degree of control and situational awareness that standard defenses cannot offer, especially once stolen data leaves its originating system. State actors must understand that the degree of invasiveness their OCM requires may produce drastically different legal and ethical issues depending if the OCM is (1) used during peacetime or during hostilities or (2) used against foreign actors or USPs.

The OCMs discussed in Section 2 present a possible evolution of digital defense techniques. How nations choose to implement such OCMs may alter worldwide perceptions of these techniques. These OCMs could represent the first viable cyber deterrent for protecting systems such as our nuclear command and control systems, or, these OCMs could be the antagonizing factor that triggers the next kinetic conflict. Tallinn 2.0 briefly discusses the use of OCMs, but there is yet to develop an international norm or binding law governing their use. Nations must determine if they will proactively recognize the set of legal and ethical issues OCMs create and codify norms for their use; alternatively, if nations maintain OCMs as a clandestine defense and deal with the ramifications after the global discovery of their use. Our discussion of controversial OCMs such as poisonous honeypots does not constitute our endorsement of those tactics but is meant to trigger follow-on discussions about its place in defending sensitive intellectual property and information.
NOTES

NOTES

25. Laurel O’Connor, “Celebrity nude photo leak: Just one more reminder that privacy does not exist online and legally, there’s not much we can do about it,” Golden Gate University School of Law Review Blog, 2014, https://goo.gl/X3b4GK.


27. Ibid., §1.8.1.

28. Ibid., §1.8.


30. In this context, CNE is the penetration into targeted digital systems for observation and gathering intelligence data.


33. Ibid., r. 97.


35. 22 U.S.C. §6010 defines a USP as “any United States citizen or alien admitted for permanent residence in the United States, and any corporation, partnership, or other organization organized under the laws of the United States.”


52. Ibid., r. 92, cmn. 10.


NOTES

69. Ibid., r. 69, cmt. 8.
72. Ibid., r. 73.
73. Ibid., r. 26.
74. Since a nation defines its own CIKR, a nation could simply define everything as CIKR and essentially authorize the use of force against other nations at will.