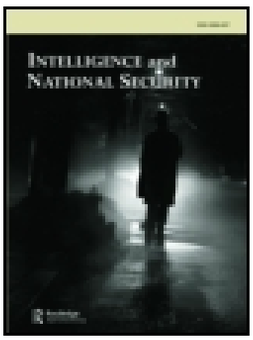

Intelligence and National Security

ISSN: 0268-4527 (Print) 1743-9019 (Online) Journal homepage: http://www.tandfonline.com/loi/fint20

Aesop's wolves: the deceptive appearance of espionage and attacks in cyberspace

Aaron F. Brantly

To cite this article: Aaron F. Brantly (2015): Aesop's wolves: the deceptive appearance of espionage and attacks in cyberspace, Intelligence and National Security, DOI: 10.1080/02684527.2015.1077620

To link to this article: http://dx.doi.org/10.1080/02684527.2015.1077620

Published online: 15 Sep 2015.

INTELLIGENCE AND NATIONAL SECURITY, 2015
http://dx.doi.org/10.1080/02684527.2015.1077620

Routledge
Taylor & Francis Group

ARTICLE

Aesop's wolves: the deceptive appearance of espionage and attacks in cyberspace

Aaron F. Brantly

ABSTRACT
Appearances in cyberspace are deceptive and problematic. Deception in the cyber domain poses an immensely difficult challenge for states to differentiate between espionage activities in cyberspace and cyber attacks. The inability to distinguish between cyber activities places US cyber infrastructure in a perilous position and increases the possibility of a disproportionate or inadequate response to cyber incidents. This paper uses case analysis to examine the characteristics associated with the tools and decisions related to cyber espionage and cyber attacks to develop a framework for distinction leveraging epidemiological models for combating disease.

The prominent debates on offensive and defensive balances of power, deterrence, and threat illustrate the importance of identifying a state's capabilities prior to crafting adequate strategic and tactical responses. The differentiation of capability in international relations is fundamental to the determination of adequate state response behaviors. Yet, unlike with nuclear and conventional weapons, the classification of capabilities in cyberspace is often conducted post-hoc. Post-hoc capability identification is problematic both linguistically, in the way in which the discourse in the field of cybersecurity is conducted, and technically, in the way in which cyber weapon systems (CWS) are created and utilized. The proper classification of accidents, attacks, criminal behaviors, or espionage activities in cyberspace is presently conducted through forensic analysis and contextual relationships to current or recent events. At present, the tools and techniques in place to quickly and accurately classify CWS operate below necessary operational tempos and result in delays that limit the effectiveness of state strategic, tactical, and operational responses. Tailored, adequate responses are unable to be engaged in until the intent of a weapon system is known. This paper establishes a model to remedy the time gap between attack and response within cyberspace that is relevant and suited to intelligence practitioners.

The modern vernacular associated with CWS is largely derived from a biological context that leverages the terms virus, trojan, and worm, among others. States largely use post-hoc analytical structures for the classification of capabilities. Understanding the time constrains placed upon modern cyber operators in the Department of Defense, is it possible to leverage existing analytical models from other fields of study to differentiate, assess, and treat cyber-related incidents? Specifically, are there non-Information Communications Technologies frameworks better suited to the analysis of CWS than post-hoc forensic analysis?

This work leverages both international laws analyzed in the *Tallinn Manual* and aspects of epidemiology (EPI) and biology leveraging the concepts of basic incidence, prevalence, case fatality, and

CONTACT Aaron F. Brantly aaron.brantly@usma.edu

reproduction number. The purpose of this work is to further discussion within the intelligence and operational cyber community on the development, assessment, and identification of CWS for the purposes of offensive cyber operations (OCO) and defensive cyber operations (DCO). The development of a framework for CWS differentiation provides a mechanism to avoid conflict escalation similar to how doctors select the appropriate drug to treat a sick patient. The clinically informed assessment of a cyber incident can facilitate more accurate and timely mitigation of the disease and foster future patient resilience or the cyber equivalent of confidentiality, integrity, and availability (CIA Triad) of assets and networks. It should be noted at the outset that this is a general proof of concept and prone to weaknesses, which will be discussed in the conclusion.

As intelligence becomes increasingly cyber enabled, there is an ever growing need to develop novel analytical frameworks for the collection and analysis of threats to national security emanating from cyberspace. Not only is it important to leverage existing SIGINT, MASINT, GEOINT, HUMINT, and others, scholars and practitioners necessarily need to address CYBERINT. CYBERINT can provide intelligence in support of within domain and cross-domain operations. This paper provides a step in the direction of establishing a framework within which to determine the behavioral attributes of actions occurring within cyberspace.

Assessing attributes of cyber weapons legally

Determining the attributes of behavior in cyberspace is problematic in many, but not all, instances. Cyber weapon systems are tacitly considered offensive and yet their fundamental distinguishing factor is not necessarily the code itself, but rather the intent associated with the use of that code on a target system. A given malware or malicious behavior resulting in cyber espionage (CE), offensive cyber operations, or defensive cyber operations-response action (DCO-RA) is often indistinguishable absent context or significant forensic analysis. The absence of context places cybersecurity within the doubly dangerous environment explained by Robert Jervis in which offense has the advantage and offensive and defensive weapons are indistinguishable.[1] The advantage and indistinguishable nature of cybersecurity occurs in what Clark and Konrad refer to as an asymmetric environment in which states are defending an almost infinite number of unknown weak links against a single unknown best shot.[2]

Although states face a doubly dangerous environment in cyberspace due to a lack of distinguishing features between offensive and defensive capabilities, the relative danger of this environment is not comparable to weapons capabilities such as nuclear, biological, or chemical. The lack of cross-weapon type comparability is due to the scale and complexity of the threats in cyberspace that range from annoyances up to the potential for nuclear incidents. The argument for the distinguishability of offensive and defensive forces is commonly examined in the context of nuclear arms thus raising the profile of the doubly dangerous nature of an indistinguishable environment. However, to date, the same level of threat has not been fully demonstrated or articulated in cyberspace. Simply put, the effects of nuclear weapons are clear and the effects of cyber weapons are not, thus the doubly dangerous mantra is not meant to imply a level of violence not yet demonstrated, but rather a conceptual threat associated with the indistinguishable nature of capabilities in function from the perspective of a potential target.

For the purposes of this paper, the concern is not with offense-defense theories predictions of states in the international system engaging in a perpetual arms race as predicted on the security dilemma. Instead, the purpose of distinguishability is helpful for both the operator and the policy-maker to determine the immediate response either politically or operationally. Determining the appropriate response requires understanding cyber operations, espionage, and other behaviors within a broader

[1] R. Jervis, 'Cooperation Under the Security Dilemma', *World Politics* 30/2 (1978) pp.167–214.
[2] D.J. Clark and K.A. Konrad, 'Asymmetric Conflict: Weakest Link against Best Shot', *Journal of Conflict Resolution* 51 (2007) pp.457–69. Clark and Konrad illustrate the offensive advantage in clearly distinguishable terms using economic models and game theory.

legal framework. To date the most robust analysis of cyber operations can be found in the *Tallinn Manual on the International Law Applicable to Cyber Warfare*.[3] Rule 6 of the *Tallinn Manual* states:

> A State bears international legal responsibility for a cyber operation attributable to it and which constitutes a breach of an international obligation.

The legal responsibility of a state for a cyber operation attributable to it in violation of international obligations is a high bar to set and is rife with potential areas of misunderstanding and interpretation. Determining responsibility necessarily requires the adequate differentiation between state, proxy, non-proxy, and trans-territorial operations, all of which are present within the cyber domain. The level of obfuscation available to state actors makes the applicability of Rule 6 complex and prone to inaccuracies when examined from the perspective of weapons development or from a post-hoc forensic analysis. Embedded in Rule 6 is a degree of nuance associated with adhering to international obligations. Rules 10 and 11 of the *Tallinn Manual* parse out the nuance of the use of force directly in terms of the 'purposes' of the United Nations and further isolates the concept down to the use of a CWS that results in a scale and effect comparable to non-cyber alternatives. Diving more deeply into Rule 11, the use of force is distinguished based on historical precedent qualifying as 'armed attack'. The demarcation line as to what constitutes an armed attack is a point of contention within the broader cyber literature. Thomas Rid is the foremost proponent of the logic that most cyber operations do not come anywhere near cyber war, but are instead more closely associated with acts of sabotage.[4]

Further isolating the complexity of cyber operations is Rule 48 – Weapons Review – from the *Tallinn Manual* which says:[5]

(a) All States are required to ensure that the cyber means of warfare that they acquire or use comply with the rules of the law of armed conflict that bind the State.
(b) States that are Party to Additional Protocol I are required in the study, development, acquisition, or adoption of a new means or method of cyber warfare to determine whether its employment would, in some or all circumstances, be prohibited by that Protocol or by any other rule of international law applicable to that State.

The rule, as written, identifies a fundamental problem associated with CWS development and use. The intent of a weapon devised for one purpose can result in a use case for an entirely different purpose under differing circumstances. While the initial intent of a CWS might conform to international law or the laws of armed conflict (LOAC), a tool designed for espionage can provide the necessary access needed to conduct offensive operations. Furthermore, many of the most valuable CWS take advantage of vulnerabilities resident within existing systems. Does the acquisition of a zero-day exploit also require a weapons review? It constitutes a new means or method to engage in a broad range of cyber operations, but the vulnerability was not developed in isolation and was instead taken advantage of. Often such vulnerabilities are bought in bulk to provide future opportunities should a need arise to use them.[6] At the time of their purchase, vulnerabilities likely do not have a direct means or method attack associated with them. Just as the discovery of a genetic defect might make a given population more vulnerable to certain biological weapons does not indicate that weapons are going to be developed in response to the discovery of all genetic defects.

Kenneth Geers proposed the concept of a Cyber Weapons Convention (CWC) based on the Chemical Weapons (CW) conventions.[7] Despite developing the concept for a convention on monitoring and assessing CWS, Geers finds there are significant problems facing successful technical implementation. Geers finds that the speed of actions taken in cyberspace likely necessitate that any institutional

[3] Michael N. Schmitt and NATO Cooperative Cyber Defence Centre of Excellence, *Tallinn Manual on the International Law Applicable to Cyber Warfare: Prepared by the International Group of Experts at the Invitation of the NATO Cooperative Cyber Defence Centre of Excellence* (New York: Cambridge University Press 2013).
[4] Thomas Rid, 'Cyber War Will Not Take Place', *Journal of Strategic Studies* 35 (2012) pp.5–32.
[5] Schmitt and NATO Cooperative Cyber Defence Centre of Excellence, *Tallinn Manual on the International Law Applicable to Cyber Warfare*, p.153.
[6] Shane Harris, *@War: The Rise of the Military-Internet Complex* (Boston: Houghton Mifflin Harcourt 2014) p.96.
[7] Kenneth Geers, 'Cyber Weapons Convention', *Computer Law and Security Review* 26/5 (2010) pp.547–51.

framework or convention rely on real-time network analysis capabilities. Such capabilities are constitutive of providing weapon intent. While his concept is robust it fails to broach the issue of level of weapon used. Whereas the use of almost all chemical weapons are prohibited under CW conventions, much of what constitutes a cyber weapon falls outside of conventional weapons monitoring regimes. A country that develops sarin gas recognizes its use is far more valuable as a weapon than as a pesticide; the same distinction cannot be made for all or even most cyber weapon systems. Because the intent of a CWS weapon is not evident in its construction, cyber inspectors to any regime necessarily need to contextualize its behavior lest such a regime attempt to prevent espionage or criminal behaviors.

Law and policy in cyberspace need better mechanisms for determining intent. If it is only possible to differentiate violations by comparing them to the effects of operations in conventional domains, then the expectation is that virtual interactions or relationships have physical world corollaries. The result is the application of policy and law via analogy to a domain that derives its symbolism and its value as much from its ability to manage kinetic interactions as it does in its ability to enable dynamic, interconnected information environments through the maintenance of confidentiality, integrity, and availability.

Rule 11(9)(a-h) provides the formation of a legal and policy rubric in which a systematic understanding of actions in cyberspace can be formally examined. To differentiate whether something constitutes the use of force and is therefore an offensive weapon as opposed to a tool of espionage or other behaviors necessitates a comprehensive understanding of severity, immediacy, directness, invasiveness, measurability of effects, military character, state involvement, and presumptive legality.[8] Accurate differentiation and clarity on weapon use helps in identifying weapon type.

Unlike in conventional kinetic warfare where the use of a smart bomb has a clear intended objective, the use of targeted code might not have the same level of clarity associated with its use. Beyond a lack of clarity is a level of uncertainty associated with intent. A bomb explodes, a gun fires, a cyber weapon does what? A given CWS might not have intended to crash a network or to manipulate a programmable logic controller (PLC) for a subway system, but because of a mistake in the code it spread far beyond its initial goals. Does a malicious code, such as Stuxnet, that extends beyond its intended target warrant a military response from all nations whose systems were invaded? Breaking down the attributes of cyber weapons in a doubly dangerous environment, an environment in which offense has the advantage and offensive and defensive weapons are difficult to differentiate, requires more than understanding the use of force. It requires understanding the intent of a weapon.

An errant bomb might stray within a reasonable radius of intended target, yet an errant cyber weapon could stray to dozens of countries. The possibility for escalation might be great and unintended even if all best efforts to create a targeted weapon have been made. In cyberspace, differentiation between weapons will grow in importance as the value of cyberspace increases relative to the connections and controls it facilitates. Understanding the intended target and associated intent for damage of a weapon, not just the actual effects, helps to qualify proportionate responses. This logic differs from *Tallinn Manual* Rule 13(11) as supported by the majority of the group of experts. If intent doesn't matter, then the errant bomb that struck the Chinese embassy in Belgrade might have warranted a significant military response by China.[9] However, the intent of the act was lacking and the two governments agreed upon alternative solutions proportionate to the intent (or lack thereof). The same is true in relation to acts of violence. If only scale and effects mattered to western nations the impact of the moderately sized terror attacks in Paris in 2015[10] would have been less than those associated the large-scale terrorist attacks in Nigeria occurring at the approximately the same time.[11] Intent is an important aspect in most criminal and civil cases in domestic law and it is equally important in international law.

[8] Schmitt and NATO Cooperative Cyber Defence Centre of Excellence, *Tallinn Manual on the International Law Applicable to Cyber Warfare*, p.49.

[9] Brent Sadler, John Raedler and Carl Rochelle, 'NATO Expresses Regret, Resolve after Bombing Chinese Embassy' <http://www.cnn.com/WORLD/europe/9905/08/kosovo.03/>

[10] Mohamed Madi, Sherie Ryder, Julia Macfarlane, Alastair Beach and Victoria Park, 'As It Happened: Charlie Hebdo Attack' <http://www.bbc.com/news/live/world-europe-30710777>

[11] Aminu Abuakar and Faith Karimi, '2,000 Feared Killed in "Deadliest" Boko Haram Attack in Nigeria' <http://www.cnn.com/2015/01/09/africa/boko-haram-violence/>

In brief, law and policy necessitate the ability to distinguish between offensive actions and other forms of action, not only in the resultant effects, but also in the intent associated with the use of a given weapon. Moreover, because CWS might be qualitatively similar in all aspects within a block of code with the exception of a single logic statement changing from true to false, the resultant effect of changing the logic from collection to destruction alters the intent and shifts from espionage to attack. The use of an epidemiological model is designed to help develop a more refined sense of the scale and effect of the use of a given malware as well as the intent behind its use.

Cyber epidemiology building blocks

The level of analysis in this section resides at the national level. Locating the level analysis establishes who has responsibility for response behaviors internal to the state and the development of strategies and tactics necessary for the implementation of external responses. State level of analysis is informed with the inclusion of network level attributes and response behaviors organized within an epidemiological framework. The resultant analysis leverages attributes relevant to both malware and malicious behavior as well as to public health responses to a complex multiple pathogen environment. The concept of analyzing cybersecurity using an EPI framework is not entirely novel and has been conducted with emphasis on individual aspects of tracking and assessing the spread of contagions.[12] Santiago Gil and Alexander Knott note that EPI modeling allows for the consideration of frequencies of infection, behavioral, environmental, and other exogenous factors beyond the technical facts of a network.[13] Through the expansion of the conceptual foundation of cybersecurity outwards they demonstrate in a limited sample reliable predictive capability for enhancing cybersecurity. Cyber bugs, like their biological counterparts, can be extremely nasty, yet they display behavioral patterns and interact within environments differently.

The term malware is encompassing of a diverse set of hostile and intrusive coding schema and techniques used to adversely affect target systems. The comparative EPI term is pathogen, defined as an infectious agent that causes illness or disease in a host. Malware, as a pathogen, seeks to infect or cause disease in a host computer, system, or process. Just as not all pathogens result in death, not all malware has the same goal. The primary distinguishing factor between pathogens and malware resides in the level of control associated use. Designer pathogens can target specific biological/genetic attributes or achieve specified goals in biological warfare or can be designed to evolve within the target host to achieve novel effects. Similarly malware can act as a fire-and-forget targeted pathogen or it can be manipulated once in the host to accomplish objectives that might extend its original purpose. However, the author at present knows of no pathogen synthetic or organic that can be altered at will once within the target other than to follow evolutionary/adaptive processes. The same attributes cannot be said for malware. While malware can be single purpose or adaptive through quasi-evolutionary (adaptive or automated) processes, it can also be remote controlled to facilitate novel changes once in the target system at the request of its implementer.

Incidence and prevalence

Any discussion on epidemics necessarily starts by examining concepts of incidence and prevalence across network types and discusses the logic of point source versus propagated (progressive) epidemics. These basic frameworks help isolate the broader nature of the problem before addressing the specific attributes associated with a pathogen vector or categories of vectors. Within cyberspace incidence rate is the ratio of cases from network live time or simply the time at risk for infection. So long as a network is active it is considered live. Using the military's Nonsecure Internet Protocol Router network

[12] Santiago Gil, Alexander Kott and Albert-László Barabási, 'A Genetic Epidemiology Approach to Cyber-Security', *Scientific Reports* 4 (2014); Angel Stanoev, Daniel Trpevski and Ljupco Kocarev, 'Modeling the Spread of Multiple Concurrent Contagions on Networks', *PLoS ONE* 9/6 (2014).

[13] Gil et al., 'A Genetic Epidemiology Approach to Cyber-Security'.

(NIPRnet) as an example, incidence rate represents the number of pathogens (malware) as a percentage of network connected devices infected relative to the network live time and devices measured. Because devices are constantly entering and leaving the network, the incidence rate of a sample will experience what is referred to as exponential decay.[14] Exponential decay refers to the loss of systems over time due to all potential factors, and changes the decline in assets from one that is linear to one that declines over time. A sample of a population of computers comprised of windows machines with similar physical attributes in RAM and processor type will not all achieve the same product lifespan. Instead, some systems will be more heavily used than others and might fall out of the sample earlier, while others will likely experience infection and be removed from the sample. Therefore, although the rate over time might be consistent, the expected decline in actual networked assets within the sample is not. If a hypothetical sample from NIPRnet was comprised of 1000 computers and 100 systems fall victim to any type of malware over the course of the year then the incident rate of the sample is .1 or an expected 10% of all networked assets are likely experience a new infection in a given year.

Network administrators commonly monitor the volume of (i.e. how many) networked assets compromised. Instead, measuring the incidence rate of computers contracting malware shifts the conversation from absolute protection with zero-risk tolerance to one of risk mitigation. This moves the defensive operation from merely preventing infections to minimizing the incidence of infections. Understanding the incidence rate broadly within the network as well as nuanced rates associated with specific malware types or families helps inform decision-makers as to the level of risk faced by a given network and alleviate misunderstandings or cognitive biases based on volume of infected asset reporting. A large incidence rate in relation to other factors indicates a wider attack warranting a more serious response. However, in determining response it is not possible to rely solely on incidence rates as will be explained in more detail below.

Similar to incidence rates it is helpful to examine prevalence of infected assets within a networked environment. Prevalence is the proportion of cases of malware-infected assets present within a network at any given time (point prevalence) or over a period of time. Using prevalence, it is possible to further determine how severe a given network compromise is or how severely compromised a network is over a period of time. Both measures change the conversation to risk mitigation and inform policy-makers as to the true nature of an event occurring in cyberspace. A malware vector with a high incidence rate and a high prevalence proportion, depending on the characteristics of the malicious code, indicate serious problems necessitating a quick response by DODIN-ops or a DCO-RA team. Identification helps combatant commanders more efficiently allocate limited cyber resources to maintain broader network operations in environments where network shutdowns are simply not a viable option.

As an example, the prevalence proportion of the Morris worm reached a peak proportion at approximately 6000 computers in 1988.[15] The prevalence of the worm prior to corrective measures was ≈10%. This means that ≈10% of all computers on the Internet (the measured network) were infected at one point in time. Note this does not indicate the probability of infection only the volume of compromise associated with the network. In the months prior to launch the incidence rate of the worm was 0%. To meaningfully measure the incidence or the probability of infection it is necessary to track the infection rate over time and provide a measure of that rate. Assuming the network size was approximately 60,000 Internet connected computers and the timeframe for measure was one week. The incidence rate indicates that networked assets had a 10% likelihood of becoming infected over the course of the week barring corrective action or risk mitigation. Because networks like human communities are dynamic the incidence rate changes as the population changes. The shortened time horizon of a week indicates that the probability of contracting the Morris worm during the initial week was ≈10%. The point prevalence proportion of the Morris worm was extremely high in the first week. As the worm made its way across networks and computers were taken offline or the worm was removed, the prevalence proportion declined but the incidence rate remained high until sufficient steps were taken to mitigate the risk to the remaining systems.

[14] Kenneth J. Rothman, *Epidemiology: An Introduction* (NY: Oxford University Press 2012) p.1196.
[15] Larry Seltzer, 'The Morris Worm: Internet Malware Turns 25' <http://www.zdnet.com/article/the-morris-worm-internet-malware-turns-25/>

As malware evolves it becomes necessary to assess why a type of malware behaves the way it does. A targeted CWS will likely have an extremely high incidence rate as it moves from system to system to find its intended target and yet the point prevalence proportion on any given day or over a period of time might be low. Had Stuxnet contained a self-delete function in its final iterations this model would have accounted for such a difference. Stuxnet was adept at infecting new machines and seeking out specific attributes.[16] It was also likely a mistake that it did not contain a self-delete function only activated in the presence of specific attributes. If a self-delete function had been operational, Stuxnet would have demonstrated high incidence and low prevalence. Malware with a high incidence rate and low prevalence rates not attributable to DCO mitigation efforts are likely indicative of targeted CWS. Similarly, malware with high prevalence and low incidence are likely indicative of advanced persistent threats, as the malware is highly pervasive and resilient in the face of potential DCO mitigation strategies, yet the risk of any new system diagnosis is low.[17]

Case fatality and reproduction numbers

By isolating the broad aspects of an epidemiological framework for cyber weapon systems analysis, the model begins to shed light on to the behavior and intent of a given CWS. Yet differentiation of capability requires still further nuance. Malware results in different effects on different computers and networks based on a variety of factors. Another important tool is a measure that provides insight into the severity of a malware epidemic, known as the case fatality rate. Unlike in biology, the case fatality rate does not necessarily constitute the 'killing' of a computer system; in cyberspace it reflects a spectrum of outcomes tailored to the functions of network. A case fatality for most critical infrastructure systems is likely constitutive of the compromise of any aspect of the CIA triad. The lower end of a hypothesized case fatality spectrum consists of violations of confidentiality case fatalities of confidentiality are more likely associated with espionage or other non-offensive behaviors. Violations of the integrity or the availability of the system likely indicates a system should no longer be relied upon and is therefore the equivalent of a fatality. Moreover, such a compromise is indicative of offensive behavior.

The case fatality rate is the proportion of systems infected that suffer from a compromise of the CIA triad. Using case fatality, the severity of any particular piece of malware is assessed relative to the analyzed population or the effects associated with a node within a population adversely affecting critical assets which require integrity and availability (i.e. mission critical attributes). At its most basic, case fatality is representative as a measure of network resilience.

Returning again to the Stuxnet example, the main malware vector resulted in a high case fatality rate in systems with specific attributes, namely exploiting four specific Microsoft windows Zero-Day[18] vulnerabilities and specific Siemens PLC model S7-315-2 running a specific Profibus communications processor 342-5 connected to a frequency converter manufactured by one of two countries.[19] The case is an example of one in which the global population case fatality rate of the malware was quite small while the target case fatality rate is estimated to have been quite high. Here the epidemiological model provides relevant insight into the actor type of the model based on inference. Most malware is indiscriminate in nature, meaning that the population and sample case fatality rates should be similar. If samples of malware case fatality rates are taken globally the rates of most malware are likely consistent, however, because states under the Geneva Convention Article 51(5) both sections (a) and (b) provide for the discrimination of attacks between military and civilian targets. This led the authors of the *Tallinn Manual* to establish Rule 50: [20]

[16] Kim Zetter, *Countdown to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon* (New York: Crown Publishers 2014).

[17] This assumes that adversary objectives are focused on specific systems rather than on a less sophisticated 'buck shot' approach.

[18] A zero-day exploit is a previously unknown exploit.

[19] N. Falliere, L.O. Marchu and E. Chien, *W32 Stuxnet Dossier* 4 (2012) pp.1–69 <http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf>

[20] Schmitt and NATO Cooperative Cyber Defence Centre of Excellence, *Tallinn Manual on the International Law Applicable to Cyber Warfare*, p.157.

> A cyber attack that treats as a single target a number of clearly discrete cyber military objectives in cyber infrastructure primarily used for civilian purposes is prohibited if to do so would harm protected persons or objects.

An unbalanced case fatality rate across samples within a population is indicative of adherence to international law at the broadest level and provides distinction in terms of the objectives of an attack. This type of differentiation is likely to be indicative of a state based offensive behavior or state based espionage. As the malware payload is examined, the distinction between espionage and offensive actions becomes more readily apparent. What is important is that by adding case fatality to incidence and prevalence, case fatality the model further distinguishes both attack and attacker attributes.

Breaking case fatality down further for cyberspace, it is possible to isolate which aspect of the CIA triad was violated. While broadly in cyberspace a fatality is considered a violation of any aspect of the CIA triad, we can further distinguish the offensive nature by assessing whether the fatality was related to confidentiality, integrity, or availability. Arguably each of these aspects of system maintenance is important, however when distinguishing between offensive and espionage activities the violations of integrity and availability are indicative of a level of immediate severity likely affecting or liable to affect current operations and asset availability, while it is possible that a violation of confidentiality might lead to violations of integrity and availability at a later time. Therefore parsing out the type of case fatality facilitates an understanding as to whether the attack is offensive or espionage, state or non-state related. Case fatalities associated with integrity and availability of assets within the network are likely to be offensive in nature as they in some way degrade, disrupt, or deny the normal functioning of systems. On the contrary, the violation of confidentiality most clearly constitutes espionage behaviors. Unbalanced relationships are indicative of targeted behavior and likely indicative of state involvement, whereas indiscriminate behavior is more likely to indicate non-state or proxy behaviors.[21]

Having established the probability of infection, the prevalence of infection, and the case fatality of infections, the focus necessarily shifts to the infectiousness of any given type of malware. Understanding the basic reproductive number of malware facilitates a dynamic understanding of how it interacts with the environment. The basic reproductive number, R_0, is a measure of the infectiousness of a given type of pathogen or, in the case of cyberspace, malware. Determining the R_0 of malware provides additional evidence as to malware behavior, the behavior of the actors behind a given form of malware, and the response of the network itself to infection.

Research has been conducted in computer science to develop responsive network defense strategies to counter the infectiousness of various malware. Noam Goldberg, Sven Leyffer, and Illya Safro from the Argonne National Laboratory model using a deterministic network response optimization model leverage birth rate and death rate, and segment the network into three categories: infected, recovered, and susceptible.[22] Their work develops a deterministic model to automate network defense based malware spread. The work by Deepayan Chakrabarti, Yang Wang, Chenxi Wang, Jurij Leskovec, and Christos Faloutsos examines in detail the virus propagation in networks.[23] Both of these works provide insight into network defense strategies by focusing on those nodal points and the mechanisms needed to maintain network integrity, yet what they fail to do is provide insight into how a state might assess a given type of attack. These models are agnostic in terms of offensive and espionage or other behaviors.

Although they model themselves as agnostic they do provide mechanisms to assess the reproductive number of a given malware and how to mitigate that malware in a network. Taking the R_0 of a given malware, is it possible to enhance concepts of CWS differentiation? What does a high $R_0 > 1$ or a low $R_0 > 1$ number indicate in terms of CWS differentiation, and does it help the policy-maker determine extra-network response actions.

[21] More data are necessary to fully develop and hone the accuracy of the model. This assessment is based on anecdotal evidence across several dozen attacks; James A. Lewis, *Significant Cyber Incidents Since 2006* (Washington, DC: Center for Strategic and International Studies 2006) <http://csis.org/files/publication/140310_Significant_Cyber_Incidents_Since_2006.pdf>

[22] Noam Goldberg, Sven Leyffer and Illya Safro, *Optimal Response to Epidemics and Cyber Attacks in Networks* (Argonne, IL: National Laboratory 2012).

[23] Deepayan Chakrabarti, Yang Wang, Chenxi Wang, Jurij Leskovec and Christos Faloutsos, 'Epidemic Thresholds in Real Networks', *ACM Transactions on Information and System Security* 10/4 (2008) pp.1–26. doi:10.1145/1284680.1284681.

PCWorld magazine in 2011 provided a list of five 'fast' spreading malware, each of which highlights the problems faced by policy-makers and network defenders alike.[24] Zeus, Sality, Gamevance, Hotbar, and Fakesysdef each have different attributes, yet the vectors of attack for four of the five are human error in the form of clicking on malicious links or visiting infected websites. Only Sality in 2011 contained a self-replicating function to create a propagated-progressive-source epidemic in addition to a point-propagated epidemic by copying itself onto removable media. The other four were point-propagated epidemics or multi-point-propagated epidemics, meaning that a user had to visit a location to become infected. The distinction between point-propagated and propagated (progressive-source) epidemics is similar to the distinction between food poisoning and catching the flu. Both malware epidemics can be extremely damaging. However, in contrast to biological epidemic models, the case distinction is not equivalent. Whereas point source epidemics in biological models result in high incidence rates among proximate populations to the point-source, the same might not be true in cyberspace. Point-propagated cyber epidemics might not have a single spike in infections; infections might be spread over a longer temporal range creating varying rates of point prevalence. Point-propagated malware is likely to have a $R_0 > 1$, indicating that a given epidemic is likely to die out if the point-source is mitigated. If a point-source malware results in a spike in the number of infections it is probable that the point-source is located in a highly trafficked area increasing the risk to the general population.

A progressive-source or a combined point-progressive-source propagated epidemic is more likely to contain a $R_0 > 1$. A progressive-source or combined point-progressive-source epidemic makes it more difficult to eradicate the malware unless network wide actions such as fixing zero-day vulnerabilities or updating anti-virus dictionaries were completed. Does the reproductive number of a given malware offer help in distinguishing between offensive and espionage activities in cyberspace? Independent of other factors, the reproductive number associated with any given type of malware is unlikely to offer a sole measure of distinction, but it can offer insight into the target of the attacker or spy, particularly when focusing on networks that are air-gapped.[25] The level of threat associated with a progressive source malware that is able to 'jump' the air gap indicates a level of sophistication beyond what is necessary for common criminal behavior and hints at espionage or offensive cyber behaviors. Furthermore, combing theR_0 measure with incidence, prevalence, and case fatality offers additional evidence of the complexity and sophistication of the CWS package.

Building a cyber EPI model

Combining these independent measures vwe can begin to leverage the power of epidemiology to differentiate between offensive, espionage, and criminal/other activities in cyberspace. Each attribute of an epidemic provides clues as to the nature of a given piece of malware. Unlike in human epidemiology where the value of each life is approximately equal, not all assets and not all networks are equal in cyberspace. The compromise of the Secure Internet Protocol Router network or SIPRnet would result in more damage than the NIPRnet. The compromise of select nodes within each of these networks is likely to result in greater damage than other nodes. Likewise, a network at a nuclear power facility is more important than the average business network and might contain relatively fewer networked assets, making the population smaller. A combination of attributes and variables needs to be taken into consideration. Leveraging the framework in the context of the value of the network can lead to a better differentiation between CWS. Figure 1 indicates a hypothetical view in the context of a network assuming point or range estimates occur at the congruence of the relationship between the various attributes of an epidemic.

The point of congruence and changes in the point of congruence over time if a panel is sampled longitudinally provides a strong indication of the weapon capability in the form of intent. Without breaking

[24] Chandra Steele, '5 Fast-Spreading Computer Viruses', *PC*, 11 November 2011 <http://www.pcmag.com/slideshow/story/290466/5-fast-spreading-computer-viruses>

[25] Air-gapped networks are networks physically separated from all insecure networks.

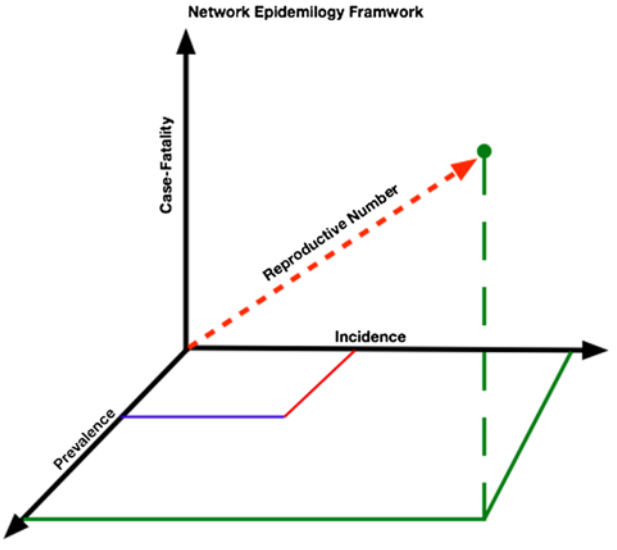

Figure 1. Hypothetical model for network epidemiology.

apart the code and only viewing the effects of a given malware the EPI model assists in isolating the behavioral attributes of a given CWS. The distinctions between offensive, espionage, and other acts lie at the congruence of the various relationships much the same way an epidemiologist puts together the pieces of an epidemic a network team can track behavior of a CWS to more accurately identify its intent.

Note that the model is network dependent. As additional networks are layered into the model the ability to differentiate attributes becomes increasingly complicated. Combining multiple networks is constitutive of the difference between within species and trans (between) species pathogens/malware.[26] Most forms of malware fall into a trans species category, yet the way in which malware affects a private network is likely to receive different attention than if it infects a military network and the malware itself is likely to behave differently in different environments. Network configurations, network assets, and attributes all differ in size, composition, and importance. Network construction establishes an important caveat in understanding distinction of capability using an epidemiological model. Distinction of capability relates to the need to isolate the network function and size in relation to an attack or incident occurring. An air-gapped network within a nuclear power facility might be very small. If it were invaded by criminal malware, the outcome might be high incidence and high prevalence and a high reproductive number, but if the case fatality remains isolated to issues of confidentiality the malware is likely related to espionage or criminal behavior rather than offensive actions. If the malware prevalence is low, the incidence high, the reproduction number is high, but the case fatality for integrity and availability is low, then the malware is likely to be a targeted CWS for espionage purposes.

Figure 2 illustrates a potential catastrophic scenario (i.e. a potential 'Cyber Pearl Harbor') in which all aspects of the epidemic are maximized. If such a malware invaded a critical network in the electric grid it would indicate an offensive weapon of significant capability, particularly if it maximized case fatality in the form of compromised integrity and/or availability.

Figure 2 is indicative of a severe malware infection resulting in catastrophic network failure. To date there have been no comparable malware equivalents. Such malware is likely to be most closely associated with pre-vaccine epidemics such as smallpox, which spread rapidly and had devastating effects. Expanding the model to a global multi-network model such as Stuxnet indicates that the CWS possessed

[26] Andy Fenton and Amy B. Pedersen, 'Community Epidemiology Framework for Classifying Disease Threats', *Emerging Infectious Diseases* 11/12 (2005) pp.1815–21.

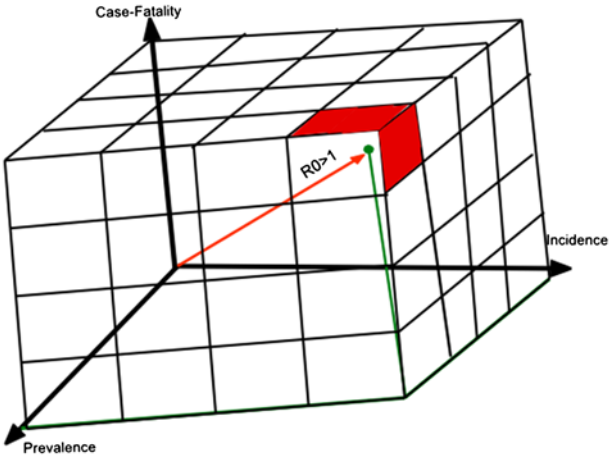

Figure 2. Severe malware infection identified using epidemiological model indicating catastrophic network failure.

limited case fatality, moderate incidence, moderate to low prevalence, and a moderate reproductive number. However, if the network EPI model is limited to Iran and specifically to the air-gapped network of Natanz, the relationship of the malware to its intended target or to the 'vulnerable' population becomes more apparent. A Natanz centric view of Stuxnet indicates a point-source-progressive malware with high incidence, high prevalence, a moderate $R_0 > 1$ and a high case fatality rate across both integrity and availability of two different categories of network assets within the network. This indicates that a larger network of networks view of malware might obfuscate or reveal its true nature. In the case of Stuxnet, the divergence in behavior between the larger population provided insight into the intent behind the malware only when deconstructed post-hoc using forensic analysis or when analyzed within the smaller network at Natanz.

Distinguishing offensive actions in cyberspace

Despite the development of an EPI framework for facilitating a more rigorous understanding of intent associating with malware in cyberspace, the problem still remains. Does the EPI framework presented above truly further the study of cyber for scholars on offense-defense theories? Initial work in this area helped by defining issues such as prevention of disease or malware and provided a far larger framework rooted within the Public Health literature.[27] David Parker and Csilla Farkas indicate: 'predictive risk would allow for the quantification of prevention interventions and allow decision-makers to see which aspects of security have the least risk'.[28] Adding to this, a more comprehensive EPI framework including network level and network of networks (trans-network EPI), a complex and dynamic picture of malware behavior and actor intent is possible although not certain.

Individuals with operational cyber experience on National Mission Teams (NMTs) indicate that the tracking of malware or malicious actor behavior within networks poses extreme difficulty until the point where CE or OCO target was reached. Operators indicate tailored access operations (TAO) engaged in by state-level actors is so well done that they are nearly untraceable. TAO operators are extremely talented, and close contact through insider threat manipulations or socially engineered access points to critical systems are unable to be assessed by this model. In much the same way the poisoning of an individual by gaining proximate access is unable to be accounted by generalized EPI models, the poisoning of networks by proximate access poses difficulties to an EPI model for cyberspace.

[27] R. David Parker and Csilla Farkas, 'Modeling Estimated Risk for Cyber Attacks : Merging Public Health and Cyber Security', *Information Assurance and Security Letters* 2 (2011) pp.32–6.
[28] Ibid., p.36.

Although, at present, the difficulty of identifying malicious behavior occurring in real-time poses significant problems for an EPI framework, particularly for high-level state behavior, the model still holds validity for a wide variety of threats and should prove valuable to network defenders and intelligence/counterintelligence operations. As threat signatures, essentially malware fingerprints, are identified and families of malware are classified, the automation of EPI modeling will facilitate a faster response framework to a large spectrum of threats. With more than 12 million malware signatures identified quarterly as of Q4 2012, McAfee estimates that there are more than 100,000 new malware samples every day and 69 new threats every second.[29] Within the diverse malware environment, Microsoft and others have narrowed down malware into categories or families with similar traits. Cisco in their 2014 Annual security report provides a breakdown of malware families including Trojan.OnlineGames, Multiplug, Syfro, Megasearch, Zeusbot, Gamevance, Blackhole, and Hupigon.[30] Cisco researchers also identified malware by category with multipurpose Trojans and iFrames.Exploits as comprising nearly 50% of total encounters.[31] Addressing malware on an attack-by-attack, virus-by-virus, penetration-by-penetration basis seems daunting. In 2010, as USCYBERCOM was being established, General Keith Alexander, in a presentation at the Center for Strategic and International Studies, estimated that the volume of probes by unauthorized users stood at approximately 250,000 times per hour or approximately 6 million times per day.[32] If even small portions of probes were successful, the resultant damage to the confidentiality, integrity, and availability of networks upon which the United States relies for national security is likely significant. Moreover, the delineation of the intent of these attacks at the outset is often difficult to discern. However, repeated analysis of attack patterns use of malware families and categories broadly might hint at techniques, tactics, and procedures of adversaries and enhance offensive capability differentiation from espionage and other capabilities.

The fruit of such an analysis will likely be to provide insight into adversary intent. As adversary intent becomes known, questions about appropriate response behaviors will be answered. The concepts of risk associated with public health and epidemiology are beneficial for creating a nuanced treatment and prevention framework. Arguably, these concepts help frame the debate over types of attack that might warrant state level responses and those that should not. Rather than disaggregated assessments of volume of attacks or even nuanced analysis of the scale and effect of a given piece of malware, understanding malware behavior provides a possible means to differentiate offensive and other activities in cyberspace and thereby separate the wolves from the sheep. Research into this space is evolving. Just as US strategy in the cyber domain is evolving, a multi-disciplinary approach to informing both technical and policy concepts is likely to provide future benefits.

The views expressed here are those of the author and do not reflect the official policy or position of the Department of the Army, Department of Defense, or the US Government.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes on contributor

Aaron F. Brantly is Assistant Professor of International Relations and Cybersecurity in the Department of Social Sciences, Cyber Policy Fellow at the Army Cyber Institute, and Cyber Policy Fellow at the Combating Terrorism Center at the United States Military Academy at West Point, USA.

[29] McAfee, 'The State of Malware 2013' <http://www.mcafee.com/us/security-awareness/articles/state-of-malware-2013.aspx>

[30] Cisco 2014 Annual Security Report <http://www.cisco.com/web/offer/gist_ty2_asset/Cisco_2014_ASR.pdf> p.38.

[31] Ibid., p.39.

[32] Sean Lawson, 'Just How Big Is The Cyber Threat To The Department Of Defense?', *Forbes*, 4 June 2010 <http://www.forbes.com/sites/firewall/2010/06/04/just-how-big-is-the-cyber-threat-to-dod/>