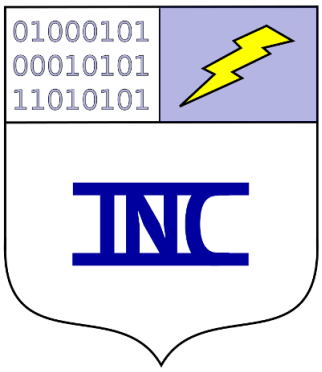
# Encouraging Vulnerability Bounties

B. Kirkpatrick *

June 26, 2016



Intrepid Net Computing




www.intrepidnetcomputing.com

*bbkirk@intrepidnetcomputing.com

## Document Revision History

**December 19, 2015** Published original on www.intrepidnetcomputing.com

**June 26, 2016** Created title page

**Abstract**

The economics of severe cyber-attacks make little sense for the attacker. This manuscript uses probability theory to demonstrate that in an adversarial setting, more intense cyber-attacks provide the defending party with more economic potential. This result indicates that sever cyber-attacks are irrational.

The results of this work suggest that industry would be well-served by offering bounties for vulnerability information. Several types of bounties are explored, including those that are correlated to the economic cost of an exploited vulnerability. This work concludes that it is sufficient to offer constant bounties, where the bounty for every vulnerability would be valued identically. Such bounties discourage potential attackers from actually carrying out attacks, as they would be giving up economic value.



# 1 Introduction

A great deal of discussion is happening in political spheres around the potential for cyber-warfare between nations. Much of this discussion remarks at the low entry cost for a nation to pursue cyber-war. There is little discussion of the economics of such a war, and this manuscript attempts to begin discussing this.

Cyber-warfare is the use of hacking and computer exploits to damage the economic potential of a rival. Most exploits are bots of one description of another, including viruses, worms, trojans, etc. Most exploits are eventually neutralized by patches and virus scanners.

Much is made of the *zero-day* period for an exploit which is the time between which the exploit is released and the appearance of the first patch. Certainly not all users will update at exactly the time the patch appears, consequently there is some variation in when updates are applied.

Certainly, the zero-day period is not the only variable in a successful defense, since an alert defending team would create *work-arounds* and infrastructure solutions that can stop attacks. For example, if a critical exploit uses open wireless networks as an attack vector, any computer that disables open wireless would successfully defend against the exploit.

Indeed these work-arounds provide essential knowledge as to the critical vulnerabilities. Finding these work-arounds gives the defending team a very good idea which part of the system is targeted by the exploit. This knowledge can lead to successful creation of a patch, sometimes without even bother with the time-intensive step of isolating the foreign computer code.

# 2 Model

Information theory was introduced by C. Shannon as he quantified the bandwidth of a communication channel. In a sense, a cyber-attack is a communication channel and the information conveyed by an attack is the vulnerabilities of the system.

We will compare the entropy of multiple cyber-attacks. This assumes that each instance of the attack uses the same exploits and involves the same amount of time under attack. We also compare the value of exploits in terms of bounties offered for their discovery.

Let there be $n$ exploits, each used by the adversary with probability $p_i$ for $0 \leq i \leq n$ where $\sum p_i = 1$. Assume that each exploit is equally effective and works all the time. Let the *de novo* probability of discovery by the defense of each exploit is $f_i$, $0 \leq i \leq n$ and $\sum f_i = 1$.

Let $s$ be the coefficient for the attack severity, where $0 \leq s \leq 1$. This means that $s = 1$ is the most severe attack. Severity roughly corresponds to the relative economic cost of the attack.

For each attack, the probability that the defending team discovers the exploit $i$ is as follows

$$\left(\frac{1}{z}\right) f_i p_i.$$

The $z = \sum_i f_i p_i$ is the partition function which makes these multinomial probabilities a distribution (i.e. the probabilities sum to one).

The defenders learn more from more obvious attacks, i.e. attacks of greater severity or greater frequency. For each exploit, the defenders learn the vulnerability information with probability given by the severity times the probability of discovery. The defenders learn about a vulnerability with probability $s$ after having discovered it. The expected fraction of vulnerabilities that the defenders identify is

$$(1/z) \sum_i s f_i p_i = s.$$

The entropy of the attack does not depend on the severity and is as follows:

$$-\sum_i (1/z) f_i p_i \log ((1/z) f_i p_i)$$

# 3 Results

Shannon entropy is a notion of the disorderedness of a communication. The entropy is greatest when there is the least information in the communication.

**Lemma 1.** *The entropy is maximized by $f_i = 1/p_i$ for all $i$.*

Two attacks can be compared based on their relative entropy which is also the Kullback-Liebler divergence. An attack that concentrates on using a single exploit would have less entropy than an attack that equally employs many exploits.

Suppose that industry offers a bounty for each vulnerability reported, regardless of whether a defender or an attacker reports it. Proprietary software vendors have offered bounties in the past for critical vulnerability information, so that they can protect their customers from exploits. Then, the economic potential of the attack for the defenders is related to the expected bounty.

Let the bounty for each exploit be proportional to the severity of the attack. The severity should roughly correlate to the cost of the attack.

$$c_i \propto s.$$

The bounty that the defense expects to get for a single exploit is proportional to the probability of that exploit's vulnerability being learned

$$(1/z) c_i s f_i p_i \propto s^2.$$

The total bounty expected for all the exploits

$$(1/z) \sum_i c_i s f_i p_i \propto s^2 n$$

is valued in the same units as $c_i$. If bounties are given in units of millions of dollars then the expected bounty is in units of a million dollars.

This model illustrates the economic reasons that most cyber-attacks go undiscovered. It is to the attacker's benefit that the defenders *do not* discover and repair each vulnerability. However, most attacks involve actions that can lead to discovery, meaning that $s > 0$.

Suppose instead of attacking, the adversary received the bounties for the vulnerabilities. Then the attacker would get a total bounty of

$$\sum_i c_i \propto sn$$

In attacking with severity $s$, the attacker gives up a portion of their possible bounty to the defense. The fraction of the bounty that the attacker retains is

$$\sum_i c_i - s^2 n = sn - s^2 n \propto (1 - s) sn$$

For models cases where $s = 1$, the attacker is penalized by loosing all the economic value of the vulnerabilities. Otherwise the attacker retains the fraction $1 - s$ of the bounty.

Under this bounty with severe attacks (large enough $s$), we see that there is no free lunch for the attacker. For the privilege of attacking, the attacker gives up the entire bounty in exchange for damage inflicted. The attacker's loss of in terms of bounties is dependent on how the bounty is set and how severe the attack is. We consider two more options for setting the bounty, both of which benefit the defense.

Suppose that the bounties not determined by the severity of the attack, meaning that the bounty is constant. In this case the attacker would retain a fraction of the bounty.

The expected bounty for the defender when $c_i = 1$ is $sn$. The expected bounty for the attacker is

$$\sum_i c_i - (sn)/z \sum_i c_i f_i p_i = n - sn = n(1 - s).$$

In this case, when the severity is $s > 1/2$, the defenders have larger economic potential.

If the bounties have constant values, then the attacker also has no free lunch. The expected bounty for $c_i = c$ constant is

$$\sum_i c_i - (sn)/z \sum_i c_i f_i p_i = \sum_i c_i(1 - sn) = nc(1 - sn).$$

This means that the attacker gives all the economic value any time that $s > 1/n$. Of the three bounties, the constant bounty most favors the defenders and most penalizes the attackers.

In practice a software vendor would probably offer a bounty exactly once, whereas this expectation computes fractional bounties. In this case, the bounties would probably further favor the defenders as they would discover the exploits and request bounties before the attackers could find out that the defenders had identified the exploits.

# 4 Conclusions

For the defending teams that are brave enough to learn from a cyber-attack, the economic potential is tremendous. This author is extremely skeptical that any rational nation would waste their computer science economic potential and their political capital on a severe attack. Severe attacks would likely be irrationally motivated.

This model and analysis suggests that industry should offer constant bounties for vulnerability information (as it is difficult to estimate the potential economic cost of an exploited vulnerability). For the most dangerous vulnerabilities, this model suggests that industry should estimate the potential cost of exploitation and offer a bounty that is proportional to the cost. Such bounties would strongly discourage attacks by giving potential adversaries the opportunity to directly benefit economically from their knowledge. In executing an attack the attacker gives up the value of the bounty to the defenders. Furthermore attackers risk international legal and political sanctions which further discourage attacks.

Open problems involve discussing differences between defenders' exploit discovery rates. Defenders that are particularly sensitive to small deviations in system behaviors would detect more exploits. In particular, such defenders would be prone to discovering zero-day exploits and could suffer more damage from each exploit while waiting for a patch.

Additional open problems include a discussion of the ethics of paying hackers vulnerability bounties if they also make money on using an exploit. Also the ethics of white-hat hacking, where former hackers turn into security experts is also open to discussion. Is there a way to pay vulnerability bounties only to the people who have not exploited a vulnerability?

The author is particularly concerned about the intersection between discrimination and defense, as discrimination is a potential motivator for irrationally severe attacks. Minority computer scientists could have higher exploit discovery rates and suffer the greater bias involved in waiting longer for a patch. Furthermore, the computing community has a history of ignoring hacking reports by minorities (based on the own author's experience), since minorities are automatically considered less competent than their majority peers. These three factors could well combine to magnify the discrimination experienced by minority computer scientists.

## Biography

Dr. Kirkpatrick has a bachelor's in computer science from Montana State University-Bozeman, a master's and a Ph.D. in computer science from the University of California, Berkeley. Dr. Kirkpatrick is an expert in deterministic and statistical computer algorithms, and his main application area is the field of computational biology, in particular genetics. However, Dr. Kirkpatrick's algorithms expertise is general to the entire field of computer science, and he has recently had cause to specialize in algorithms for computer systems.