# Security, Cyber-crime and Scale

Cormac Herley
Microsoft Research
One Microsoft Way
Redmond, WA, USA
cormac@microsoft.com

## ABSTRACT

In a traditional threat model it is necessary and sufficient to protect against all attacks. While simple, and appropriate in high-assurance settings, we show that this model does not scale and is entirely inappropriate to the financially-motivated cyber-crime that targets two billion Internet users. The attackers who prey on Internet users are very constrained. The have finite gains, non-zero costs, and must make profit in expectation. Above all their techniques must scale. This means that they must have attacks with scalable costs or efficient ways of finding viable targets in a large population. We show that many technically possible attacks are economically infeasible. We show that incorporating target selection and monetization in addition to an attacker's technical constraints offers new directions on how defense tradeoffs can be made.

## 1. INTRODUCTION

A traditional threat model, which has been with us since before the dawn of the Internet, is illustrated in Figure 1. Alice seeks to protect her resources from Mallory, who has a suite of attacks, $k = 0, 1, \cdots, Q-1$. For the moment let's just assume (unrealistically) that $Q$ is finite and all of the attacks are known to both parties.

What must Alice do to prevent Mallory gaining access? Clearly, it is *sufficient* for Alice to block all $Q$ possible attacks. If she does this, there is no risk. Further, assuming that Mallory will keep trying until he exhausts his attacks (or succeeds), it is also *necessary.* That is, *against a sufficiently motivated attacker,* it is both necessary and sufficient that Alice defend against all possible attacks. For many this is a starting point, *e.g.*, Schneider states [14] "a secure system must defend against all possible attacks – including those unknown to the defender." A popular textbook [13], calls it the Principle of Easiest Penetration: "An intruder must be expected to use any available means of penetration." An often-repeated quip from Schneier "the only secure computer in the world is unplugged, encased in concrete, and buried underground" reinforces the view.

### 1.1 How did Mallory meet Alice?

How does this scale? That is, how does this model fare if we use it for an Internet scale population, where, instead of a single Alice, we have many? We might be tempted to say, by extension, that unless each Alice($i$) blocks all $Q$ attacks then some attacker gains access. However, a moment's reflection shows that this cannot *always* be true. If there are two billion users, it is numerically impossible that each faces the "sufficiently motivated" persistent attacker that was our starting assumption: there simply aren't two billion attackers, or anything close to it. Indeed, if there were two million rather than two billion attackers (making cybercriminals about a third as plentiful as software developers worldwide) users would still outnumber attackers 1000-to-1. Clearly, the threat model shown in Figure 1 doesn't scale.

### 1.2 Sufficient $\neq$ Necessary-and-Sufficient

Thus, the threat model applies to some users and targets, but cannot apply to all. When we try to apply it to all we confuse sufficient and necessary-and-sufficient. This might appear a quibble, but the logical difference is enormous and it leads to absurdities and contradictions when applied at scale.

First, if defending against all attacks is necessary-and-sufficient, then failure to do everything is equivalent to doing nothing. Thus, the marginal benefit of almost all security measures is zero. Lampson expresses the problem succinctly [12]: "There's no resting place on the road to perfection."

Second, in a regime where everything is necessary, tradeoffs are not possible. We have no firm basis on which to make sensible claims, such as that keylogging is a bigger threat than shoulder surfing. Those who adhere to a binary model of security are unable to participate constructively in tradeoff decisions.

Third, the assumption that there is only a finite number of known attacks is clearly favorable to the defender. In general it is not possible to enumerate all possible attacks, the number grows constantly and there are very likely to be attacks unknown to the defenders. If failing

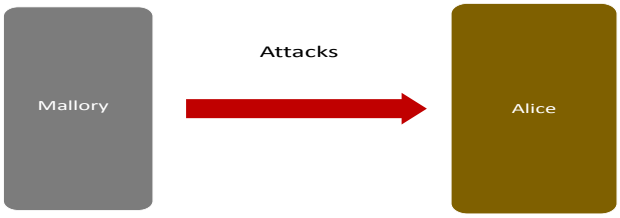


**Figure 1: In a traditional threat model a single user faces a single attacker. Given a *sufficiently motivated attacker* it is necessary and sufficient to block all attacks.**

to do everything is the same as doing nothing (and Alice can't possibly do everything) the situation appears hopeless.

Finally, the logical inconsistencies are joined by observations that clearly contradict what the model says is necessary. The fact that most users ignore most security precautions and yet escape regular harm is irreconcilable with the threat model of Figure 1. If the model applies to everyone, it is hard to explain why everyone isn't hacked every day.

### 1.3 Modifying the threat model

The threat model of Figure 1 might appear a straw man. After all, nobody seriously believes that all effort short of perfection is wasted. It's doubtful that anyone (especially those quoted above) adheres to a strictly binary view of security. Rather than insist that the threat model *always* applies many use it as a starting point that's appropriate for some situations, but is overkill for others. Thus, some modification is generally offered. The popular textbook mentioned earlier [13], for example, codifies this as the Principle of Adequate Protection "[computer items] must be protected to a degree consistent with their value." Thus, a more realistic view is that we start with some variant of the traditional threat model, *e.g.*, "it is necessary and sufficient to defend against all attacks" but then modify it in some way, *e.g.*, "defense effort should be appropriate to the assets."

However, while the first statement is absolute, and has a clear call-to-action, the qualifier is vague and imprecise. Of course we can't defend against everything, but on what basis should we decide what to neglect? It helps little to say that the traditional threat model doesn't *always* apply unless we specify when it does, and what should be used in its place when it does not. A qualifier that is just a partial and imprecise walkback of the previous claim clarifies nothing. Thus, our problem is not that anyone insists on rigid adherence to the traditional threat model, so much as that we lack clarity on when to abandon it and what to take up in its place when we do. Failure to be clear on this point is an unhandled exception in our logic.

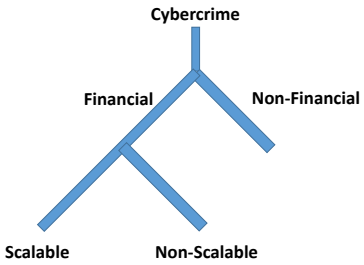


**Figure 2: Dividing attacks into financial and non-financial attacks. We further divide financial attacks into scalable and non-scalable.**

We argue that this matters. A main reason for elevated interest in computer security is the scale of the population with security needs. A main question for that population is how to get best protection for least effort. It is of the first importance to understand accurately the threats that two billion users face and how they should respond. All models may, as it is said, be wrong, but failure to scale, demands of unbounded effort and inability to handle tradeoffs are not tolerable flaws in one that seeks to address the question of Internet threats. The rest of this paper explores modifications of the traditional threat model.

## 2. FINANCIALLY-MOTIVATED CYBER-CRIME

The threat model of Figure 1 tried to abstract all context away. There is no reference to the value of the resource, the cost of the attack, or how Mallory came to focus his attention on Alice. The model doesn't distinguish between finite and infinite gain nor between zero and non-zero cost. Abstraction like this is, of course, useful. It is far more powerful if we can solve the general problem without resorting to specifics. Unfortunately, this attempt breaks down at scale: the binary view of security must be qualified.

When money is the goal it seems reasonable to assume that Mallory is "sufficiently motivated" when the expected gain from an attack exceeds the cost. We now examine whether focusing on the sub-problem of financially-motivated cybercrime will allow progress on the questions of exactly when and how to deviate from the binary model. We propose a bifurcation of attacks (shown in the top branch of Figure 2) into those that are financially motivated and those that are not.

### 2.1 Profit at scale

A main reason for the concern with cybercrime is the scale of the problem. We might be less concerned if it

were a series of one-off or isolated attacks rather than an ongoing problem. Only when the one-time costs can be amortized over many attacks does it become a sustained phenomenon that affects the large online population. To be sustainable there must first be a supply of profitable targets and a way to find them. Hence, the attacker must then do three things: he must decide who and what to attack, he must successfully attack (*i.e.*, get access to a resource) and he must monetize that access.

Clearly, a particular target isn't worthwhile if gain minus cost isn't positive: $G - C > 0$. Thus, when attacks are financially-motivated, the average gain for each attacker, $E\{G\}$, must be greater than the cost, $C$:

$$E\{G\} - C > 0. \quad (1)$$

$C$ must include all costs, including that of finding viable victims and of monetizing access to whatever resource he targets. The gain must be averaged across all attacks, not merely successful ones. Clearly, if either $E\{G\} \to \infty$ or $C = 0$ then (1) represents no constraint at all. When this happens we can revert to the traditional threat model with no need to limit its scope: Alice can neglect no defense if the asset is infinitely valuable, or attacks have no cost.

That gain is never infinite needs no demonstration. While it should be equally clear that cost is never precisely zero, it is common to treat cybercrime costs as being small enough to neglect. Against this view we offer the following arguments. First, if any attack has zero cost then all targets should be attacked continuously, and all profitable opportunities should be exhausted as soon as they appear. Instead of "why is there so much spam" we would ask "why is there so little?" as it would overwhelm all other traffic. Second, while a script may deliver victims at very low cost, the setup and infrastructure are not free. Even if we grant that a script finds dozens of victims in one day (the Internet is big after all) why should the same script find dozens more the next day, and again the day after? Why should it do so at a sustained rate? Finally, as we discuss in Section 3.3, while scripts might achieve access to resources at low cost, the task of monetizing access is generally very hard. Thus, we argue that, not only is attacker cost greater than zero, but it is the principal brake on attacker effort.

## 2.2 Attacks that scale

While we've argued that $C > 0$, it is clear that the majority of users are regularly attacked by attacks that have very low cost *per attacked user*. We'll find it useful to segment attacks by how their costs grow. Scalable attacks are one-to-many attacks which have the property that cost (per attacked user) grows slower than linearly. For example, doubling the number of users attacked, increases the cost very little [8]:

$$C(2N) \ll 2 \cdot C(N). \quad (2)$$

Many of the attacks most commonly seen on the Internet are of this type. Phishing and all attacks for which spam is the spread vector are obvious examples. Viruses and worms that spread wherever they find opportunity are others. Drive-by download attacks (where web-page visitors are attacked via browser vulnerabilities) are yet more. Non-scalable attacks are everything else. In contrast to (2) they have costs that are proportional to the number attacked: $C(N) \propto N$. We add the bifurcation, into scalable and non-scalable attacks, to Figure 2.

# 3. CONSTRAINTS ON FINANCIALLY MOTIVATED ATTACKERS

A financially-motivated attacker must decide who and what to attack, attack successfully and then monetize access. The better he can scale these activities the greater the threat that he represents to the online population. We now examine, some of the difficulties and constraints in scalable answers to these questions.

## 3.1 Scalable attacks (attack everybody)

An alternative to solving the problem of deciding who to attack is to attack everyone. Scalable attacks have inherent advantages over non-scalable ones. They reach large masses at very low cost and techniques can be propagated easily. These advantages come with severe constraints however. Scalable attacks are very visible: in reaching millions it is hard to go un-noticed. Their broadcast nature gives an alert, both to defenders and other would-be attackers. This attracts competition and increases defense effort.

Scalable attacks are a minority of attack types. It is the exception rather than the rule that costs have only weak dependence on the number attacked. Anything that can't be completely automated, or involves per-target effort is thus non-scalable as this cost violates the constraint (2). Physical side-channel attacks (which require proximity) are out, as getting close to a million users costs a lot more than getting close to one. Labor-intensive social engineering attacks (such as those described by Mitnick [10]) and the "Stuck in London" scam are non-scalable. After an initial scalable spam campaign the Nigerian 419 scam, and variants, devolves into a non-scalable effort in manipulation. Equally, spear-phishing attacks that make use of information about the target are non-scalable. While the success rate on well-researched spear-phishing attacks may be much higher than the scatter-shot (*e.g.*, "Dear Paypal customer") approaches they are non-scalable. Attacks that involve knowledge of the target are usually non-scalable. For example, guessing passwords based on knowledge of the user's dog's name, favorite sports team

or cartoon character involves significant non-scalable effort. Equally, attacks on the backup authentication questions that involve researching where a user went to highschool, *etc*, are non-scalable.

Thus, while all of us see the evidence of scalable attacks every day, it is actually a minority of of attack types that are scalable.

## 3.2 Finding viable targets

Non-scalable attacks resemble the one-on-one attacks of the traditional threat model. However, rather than an attacker who is sufficiently motivated to persist no matter what, we have one who obeys a profit constraint (1). The problem (for Mallory) is that profitability is not directly observable. It is not obvious who will succumb to most attacks, and who will prove profitable. Since $C > 0$, the cost of false positives (unprofitable targets) can entirely consume the gain from true positives. When this happens attacks that are perfectly feasible from a technical standpoint become impossible to run profitably. The cost and difficulty of deciding who to attack is almost unstudied in the security literature, however no audit of Mallory's accounts can be complete without it. Unless he has a cost-effective way of identifying targets in a large population non-scalable attacks are of little use to Mallory.

Assume that Mallory can estimate a probability, or likelihood, of profit given everything he observes about a potential target. This is the probability that the target succumbs AND access can be monetized (for greater than the average cost). Call this $P\{\text{viable}|\text{obs.}\}$. The observables might be address, zip code, occupation and any other factors likely to indicate the profitability. Without loss of generality, these can be wrapped into a single one-dimensional sufficient statistic [15]. We'll assume that the cost of gathering the observables is small relative to the cost of the attack. This makes the problem a binary classification [9], so that Receiver Operator Characteristic (ROC) curves are the natural analysis tool. The ROC curve is the graph of true positive rate, $t_p$, *vs.* false positive rate, $f_p$, an example is shown in Figure 3.

Let's examine how the binary classification constrains Mallory. Suppose, in a population of size $N$, a fraction $P\{\text{viable}\} = d$ of targets are viable. From Bayes' theorem (when $d$ is small):

$$\begin{aligned} P\{\text{viable} \mid \text{obs.}\} &= \frac{d}{d + \dfrac{P\{\text{obs.} \mid \text{non-viable}\}}{P\{\text{obs.} \mid \text{viable}\}} \cdot (1-d)} \\ &\approx d \cdot \frac{P\{\text{obs.} \mid \text{viable}\}}{P\{\text{obs.} \mid \text{non-viable}\}}. \end{aligned} \tag{3}$$

Clearly, $P\{\text{viable}|\text{obs.}\}$ is proportional to density; so, the hardness of finding a viable target gets worse as $d$ falls. A set of observables that gives a 90% chance of finding a viable target when $d = 0.01$ gives only a 0.09% chance when $d = 10^{-5}$. So, observables that promised a near "sure thing" at one density offer a worse than thousand-to-one long-shot at another.

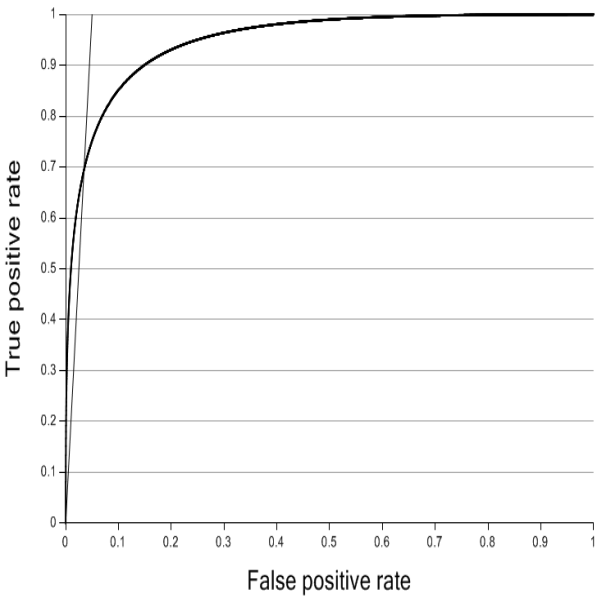


**Figure 3: Example ROC curve. A line of slope $T/d = 20$ is shown. Only operating points to the left of this line satisfy (5) and yield profit. Clearly, as $T/d$ increases the true positive rate falls and fewer viable targets are attacked. For example, with this classifier when $T/d = 10^4$ fewer than 1% of viable targets will be attacked.**

Mallory presumably decides to attack depending on whether or not $P\{\text{viable}|\text{obs.}\}$ is above or below some threshold, $T$. The threshold, $T$, will generally be set by a budget, for example if an attacker needs one attack in every $1/T$ (*e.g.*, 1-in-20, 1-in-100, *etc*) to be profitable. Then, from (3) he must have:

$$P\{\text{obs.} \mid \text{viable}\} \geq \left(\frac{T}{d}\right) \cdot P\{\text{obs.} \mid \text{non-viable}\}. \tag{4}$$

This constraint says that the observables must be a factor of $T/d$ more common among viable targets than non-viable. If 1-in-10,000 is viable and Mallory needs one attack in 20 to succeed he must identify observable features which are $T/d = 500\times$ more common in the viable population than in the non-viable.

The ROC curve gives a geometric interpretation. Mallory finds $dt_pN$ viable targets in $dt_pN + (1-d)f_pN$ attacks. To satisfy the budget constraint, the ratio of successes to attacks must be greater than $T$, so we get

(when $d$ is small):

$$\frac{t_p}{f_p} \geq \frac{T}{d}. \quad (5)$$

Thus, only points $(f_p, t_p)$ on the ROC curve to the left of a line with slope $T/d$ will satisfy Mallory's profit constraint. To illustrate a line of slope 20 is shown on Figure 3.

Since the slope of the ROC curve is monotonic [15], as we retreat to the left $t_p/f_p$ increases. Thus, (5) can almost always be satisfied for some points no matter how good or bad the classifier. However, as we retreat leftward $t_p$ decreases, so that a smaller and smaller fraction of the true positives (*i.e.*, viable targets) are attacked. For example, for the classifier in Figure 3 when $T = 1/10$ (*i.e.*, we need one attack in ten to succeed) and $d = 10^{-5}$ (*i.e.*, 1-in-100,000 is viable) we require $t_p/f_p \geq 10^4$ which happens only for values $t_p < 0.01$, (meaning that less than 1% of the *viable* population is observably profitable). As $d$ decreases Mallory ends up with a shrinking fraction of a pool that is itself shrinking [9]. Thus, without a very good classifier (which has $t_p$ high while keeping $f_p$ low), most *viable* victims escape harm.

It is easy to underestimate the difficulty of building good classifiers. Real-world examples from other domains illustrate that this is non-trivial. For example, the false positive rate for mammograms is $t_p \approx 0.94$ at $f_p \approx 0.065$ (so $t_p/f_p \approx 14.5$) [4]. For appendectomies it is $t_p \approx 0.814$ at $f_p \approx 0.105$ (so $t_p/f_p \approx 7.8$) [7]. Thus, even with the benefits of decades of effort, and millions of examples of both true and false positives, buliding a classifier is often extremely hard. This is especially true when the base-rate of sought items is low. When $d$ is small Mallory faces a seemingly intractable Catch-22: he must find victims in order to figure out how they can be found. To determine how viable and non-viable can be distinguished requires a large collection of viable targets.

### 3.3 Monetization: Access ≠ Dollars:

In many forms of non-financial cyber-crime the attacker succeeds once he gains access. Often getting the celebrity's password, control of the web-server, or the file of customer records is the end: once he's in he's done. A few screenshots, a decorated web-page or extruded files suffice if the attacker merely wants acknowledgement. However, for financially-motivated attacks things are different. The attacker isn't after passwords, or files, or access to secure servers as ends in themselves. He wants money, and is interested in these things only to the degree that they lead to money. Turning access into money is a lot harder than it looks.

For concreteness, let's consider the assets that the Internet's two billion users protect. Consider bank passwords first. It might seem that once an attacker gets a bank password that money quickly follows. However, several factors indicate that this is not the case. First, most transactions in the banking system are reversible: once fraud is discovered they are rolled back [5]. It is for this reason that many bank fraud requires money mules, who (often unwittingly) accept reversible transfers from a compromised account, and send on irreversible transfers (*e.g.*, by Western Union). A money mule can be used no more than once or twice before transactions start to bounce. Thus, while stealing passwords may be easy, and scalable, the limiting factor on the password-stealing business is mule recruitment [5]. This view also explains anecdotal accounts that the asking price for stolen credentials on underground markets is fractions of a penny on the dollar.

The situation with other account types is typically even worse. Attempts to monetize access to social networking passwords generally involve the, by now well-known, labor-intensive "Stuck in London" scam. Email accounts often receive password reset links for other accounts. However, even when a bank password can be reset, this is simply an indirect path to a resource that we already found problematic.

Other consumer assets also seem challenging. It may be possible to compromise a user's machine by getting her to click on a malicious link. However, even with arbitrary code running on the machine, monetization is far from simple. All passwords on the machine can be harvested, but we've seen that only a minority of stolen bank passwords can be monetized and most non-bank passwords are worthless. The machine can be used to send spam, but the returns on spam-based advertising campaigns are low [11]. A botnet responsible for a third of the world's spam in 2010 apparently earned it's owners $2.7 million [1]. The machine can be used to host malicious content. However, as an argument for monetization this is circular: it suggests how yet more machines can be infected, rather than how the original, or subsequent machines can be monetized. Scareware, or fake Anti-Virus appears one of the better prospects. Successfully compromised boxes can be sold: a pay-per-install market apparently pays on the order of $100 to $180 per thousand machines in developed markets [3]. Ransomware offers another possible strategy, but works best against those who do not practice good backup regimes. In summary, for a financially motivated attacker, bank passwords seem the best of the consumer-controlled assets, and that best is not very good.

Popular press accounts often paint a picture of easy billions to be made in cybercrime. However, a growing body of work contradicts this view. Widely circulated estimates of cybercrime losses turn out to be based on bad statistics and are off by orders of magnitude [6]. The most detailed examination of spam puts the global revenue earned by all spammers at tens of millions of

dollars per year [11]. In a meta-analysis of available data Anderson *et al.* [1] estimate global revenues from the stranded traveler and fake Anti-Virus scams at $10 million and $97 million respectively. The scarcity of monetization strategies is illustrated by the fact that porn-dialers (which incur high long-distance charges), which were popular in the days of dial-up modem access have resurfaced in mobile phone malware. It would be wrong to conclude that there is no money in cybercrime. It appears to be a profitable endeavor for some, but the pool of money to be shared seems much smaller than is often assumed. It is likely that those who specialize in infrastructure, and sell services on to those downstream capture much of the value.

The difficulty of monetization appears not to be clearly understood. The idea that attacks that resulted in non-financial harm might easily have been worse is quite common. The journalist Matt Honan, whose digital life was erased, but who suffered no direct financial loss, states: "Yet still I was actually quite fortunate. They could have used my e-mail accounts to gain access to my online banking, or financial services." This is almost certainly wrong. His attackers, after several hours of effort, gained access to a Twitter and an iTunes account and wiped several devices. While exceedingly inconvenient for the victim, anyone who attempts to monetize these accomplishments would likely be disappointed.

# 4. DISCUSSION

## 4.1 Scalability is not a "nice-to-have" feature

The widespread interest in computer security seems a result of scale. Scale offers several things that work in the attacker's favor. A potential victim pool that could not be imagined by the criminals of 1990 is now available. Further, a huge online population means that even attacks with very low success rates will have significant pools of victims (if only one in a million believes an offer of easy money from a Nigerian Prince, there are still 2,000 in the online population).

However, a large pool helps only if there is some way to attack it. Scalable attacks can reach vast populations, but, as we saw, they fall into a few limited categories. Non-scalable attacks face a different problem. While the number of viable victims for even a niche opportunity may be large, the hardness of finding them is related to their relative frequency not the absolute number. In this case, while the attack itself is non-scalable, Mallory still needs a low-cost way of accurately identifying the good prospects in a vast population.

## 4.2 Tradeoffs are not optional

When resources are finite, the question is not whether tradeoffs will be made but how. For defenders, a main problem with the traditional threat model is that it offers no guidance whatever on how this can be done. Most acknowledge that defending against everything is neither possible nor appropriate. Yet, without a way to decide which attacks to neglect defensive effort will be assigned haphazardly.

We are unlikely to be able to defeat unconstrained attackers who [13] "can (and will) use any means they can" with bounded effort. Recall, however, that most assets escape exploitation not because they are impregnable, but because they are not targeted. This happens not at random, but predictably when the expected monetization value is less than the cost of the attack. We propose that understanding target selection and monetization constraints is necessary if we are to make the unavoidable tradeoffs in a systematic way.

## 4.3 Which attacks can we neglect?

As before, we concentrate on attacks that are financially motivated: expected gain is greater than cost. Scalable attacks represent an easy case. Their ability to reach vast populations means that no-one is unaffected. They leave a large footprint, so they are not hard to detect and there is seldom much mystery as to whether an attack is scalable or not. In the question of tradeoffs it is hard to make the case that scalable attacks are good candidates to be ignored. Fortunately they fall into a small number of types and have serious restrictions, as we saw in Section 3.1. Everyone needs to defend against them.

Non-scalable attacks thus present our opportunity: it is here that we must look for candidates to ignore. Cyber-criminals probably do most damage with attacks that they can repeat, and for which they can reliably find and monetize targets. We suggest probable harm to the population as a basis for prioritizing attacks.

First, attacks where viable and non-viable targets cannot be distinguished pose least economic threat. If viability is entirely unobservable then Mallory can do no better than attack at random. Second, when the density of viable victims is very small $T/d$ becomes very large and the fraction of the viable population that is attacked shrinks to nothing (*i.e.*, $t_p \to 0$). This suggests that non-scalable attacks with low densities are smaller threats than those where it is high. Finally, the harder an attack is to monetize the smaller the threat it poses.

Examples of attacks with low densities might be physical side-channel attacks which allow an attacker in close proximity to the target to shoulder surf, spy on the output on a screen or printer, the input to a keyboard, and so on. The viable target density would be the fraction of all LCD screens, printers, keyboards, *etc*, whose output (or input) can be successfully attacked and monetized for greater than the cost of the attack. It seems safe to say that this fraction should be very small (perhaps

$d = 10^{-5}$ or so). It is also unclear how they might be identified. Hence, an attacker who needs one success in every 20 attacks must operate to the left of a line with slope $T/d = 5,000$ on the ROC curve. Those who can accomplish this might consider abandoning cybercrime and trying Information Retrieval and Machine Learning. Examples of resources that are hard to monetize are low-value assets such as email and social-networking accounts, etc; while these occasionally lead to gain, the average value appears quite low.

Analysis of the observability, density and monetization of attacks won't ever be perfect. To some degree judgements must be retroactive. That errors will be made seems unavoidable; however, since we can't defend against everything, attacks for which the evidence of success is clear must take priority over those for which it is not. When categories of targets (*e.g.*, small businesses in the US) or categories of attacks (*e.g.*, spear phishing emails) are clearly being profitably exploited then additional counter-measures are certainly warranted.

### 4.4 What should we do differently?

There are also possible directions for research. The hardness of the binary classification problem suggests unexplored defense mechanisms. Any linear cost component will make it impossible to satisfy (2). Imposing a small charge has been suggested as a means to combatting spam [2], and it is worth considering whether it might be applicable to other scalable attacks. Address Space Layout Randomization (ASLR) similarly converts scalable attacks into non-scalable.

Relatively unexplored is the question of making the classification problem even harder. That is, Mallory has a great sensitivity to the density of viable targets. By creating phantom targets that look plausibly viable, but which in fact are not, we make his problem even harder. For example, phantom online banking accounts that do nothing but consume attacker effort, might reduce the profitability of brute-forcing. When non-viable targets reply to scam emails it reduces return and makes it harder to make a profit [9].

We have repeatedly stressed that an attacker must choose targets, successfully attack and then monetize his success. The second of these problems has dominated research effort. However, if the admonition to "think like an attacker" is not to be empty we should pay equal attention to how attackers can select targets and monetize resources. We've pointed out the theoretical difficulty of the binary classification problem that target selection represents. Yet, for a profit-seeking attacker the problem is not abstract. It's not enough to hopefully suggest that some zip-codes, or employers, or professions might be indicative of greater viability than others. The attacker needs concrete observable features which he uses to estimate viability. If he doesn't get it right often enough, *i.e.* doesn't satisfy (4), he makes a loss. What is observable to attackers about the population is also observable to us. The problem of how viable niches for a particular attack can be identified is worth serious research. If they can be identified, members of these niches (rather than the whole population) are those who must invest extra in the defense. If they cannot it is hard to justify spending on defense. We reiterate that we have focused on financially-motivated attacks. An interesting research question would be to explore which types of target are most at risk of non-financial attacks.

## 5. CONCLUSION

When we ignore attacker constraints, we make things harder than they need be for defenders. This is a luxury we cannot afford. The view of the world where every target must block every attack is clearly wasteful and most of us understand that it is neither possible nor necessary. Yet, acknowledging this fact is helpful only if we are clear about which attacks can be neglected. The contradiction between the traditional model, which says that tradeoffs aren't possible, and reality, which says they are necessary, must be resolved. We propose that the difficulties of profitably finding targets and monetizing them are under-utilized tools in the effort to help users avoid harm.

**Acknowledgements:** the author would like to thank Rainer Böhme, Joe Bonneau, Shuo Chen, Alain Forget, Rob Reeder, Adam Shostack and the anonymous reviewers for comments that greatly improved the paper.

## 6. REFERENCES

[1] Anderson, Ross and Barton, Chris and Böhme, Rainer and Clayton, Richard and van Eeten, Michel JG and Levi, Michael and Moore, Tyler and Savage, Stefan. Measuring the cost of cybercrime. WEIS, 2012, Berlin.

[2] C. Dwork and M. Naor. Pricing via Processing or Combatting Junk Mail. Crypto, 1992.

[3] J. Caballero, C. Grier, C. Kreibich, and V. Paxson. Measuring pay-per-install: The commoditization of malware distribution. In USENIX Security Symposium, 2011.

[4] J. G. Elmore, M. B. Barton, V. M. Moceri, S. Polk, P. J. Arena, and S. W. Fletcher. Ten-year risk of false positive screening mammograms and clinical breast examinations. New England Journal of Medicine, 338(16):1089–1096, 1998.

[5] D. Florêncio and C. Herley. Is Everything We Know About Password-stealing Wrong? IEEE Security & Privacy Magazine. Nov. 2012.

[6] D. Florêncio and C. Herley. Sex, Lies and Cyber-crime Surveys. WEIS, 2011, Fairfax.

[7] L. Graff, J. Russell, J. Seashore, J. Tate, A. Elwell, M. Prete, M. Werdmann, R. Maag, C. Krivenko, and M. Radford. False-negative and false-positive errors in abdominal pain evaluation failure to diagnose acute appendicitis and unnecessary surgery. Academic Emergency Medicine, 7(11):1244–1255, 2000.

[8] C. Herley. The Plight of the Targeted Attacker in a World of Scale. WEIS 2010, Boston.

[9] C. Herley. Why do Nigerian Scammers say they are from Nigeria? WEIS 2012, Berlin.

[10] K. Mitnick and W.L. Simon. The Art of Deception: Controlling the Human Element of Security. Wiley, 2003.

[11] C. Kanich, N. Weaver, D. McCoy, T. Halvorson, C. Kreibich, K. Levchenko, V. Paxson, G. M. Voelker, and S. Savage. Show me the money: Characterizing spam-advertised revenue. In USENIX Security Symposium, 2011.

[12] B. Lampson. Usable security: how to get it. Communications of the ACM, 52(11):25–27, 2009.

[13] C. P. Pfleeger and S. L. Pfleeger. Security in computing. Prentice Hall Professional, 2003.

[14] F. Schneider. Blueprint for a science of cybersecurity. 2011.

[15] H. L. van Trees. Detection, Estimation and Modulation Theory: Part I. Wiley, 1968.