contributed articles

DOI:10.1145/2654847

Defense begins by identifying the targets likely to yield the greatest reward for an attacker's investment.

BY CORMAC HERLEY

Security, Cybercrime, and Scale

A TRADITIONAL THREAT model has been with us since before the dawn of the Internet (see Figure 1). Alice seeks to protect her resources from Mallory, who has a suite of attacks, $k = 0; 1, \ldots ,Q - 1$; now assume, for the moment, (unrealistically) that Q is finite and all 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, there is no risk. Further, assuming 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; for example, Schneider[14] says, "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" whereby "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.

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, there are many? We might be tempted to say, by extension, that unless each Alice blocks all Q attacks, then some attacker would gain access. However, a moment's reflection shows this cannot always be true. If there are two billion users, it is numerically impossible that each would face the "sufficiently motivated" persistent attacker—our starting assumption; there simply are not two billion attackers or anything close to it. Indeed, if there were two million rather than two billion attackers (making cybercriminals approximately one-third as plentiful as software developers worldwide) users would still outnumber attackers 1,000 to one. Clearly, the threat model in Figure 1 does not scale.

Sufficient ≠ necessary-and-sufficient. 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 leads to absurdities and contradictions when applied at scale.

First, if defending against all attacks is necessary and sufficient, then failure

» key insights

- A financially motivated attacker faces a severe constraint unrelated to his technical abilities; the average gain minus average cost of an attack must be positive.
- Without a cost-effective way of telling profitable targets from unprofitable targets, an attack is no use to a financially motivated attacker.
- The difficulty of finding profitable targets is extreme when density is small; a 10x reduction in the density of profitable targets generally results in much more than 10x reduction in economic opportunity for the attacker.

ILLUSTRATION BY DANIEL HERTZBERG

to do everything is equivalent to doing nothing. The marginal benefit of almost all security measures is thus zero. Lampson[11] expressed it succinctly: "There's no resting place on the road to perfection."

Second, in a regime where everything is necessary, trade-offs are not possible. We have no firm basis on which to make sensible claims (such as keylogging is a bigger threat than shoulder surfing). Those who adhere to a binary model of security are unable to participate constructively in trade-off 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, as the number is growing constantly, and there are very likely to be attacks unknown to the defenders. If failing to do everything is the same as doing nothing (and Alice cannot 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 difficult to explain why everyone is not hacked every day.

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

However, while the first statement is absolute, and involves a clear call to action, the qualifier is vague and imprecise. Of course we cannot defend against everything, but on what basis should we decide what to neglect? It helps little to say the traditional threat model does not 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 original claim clarifies nothing. Our problem is not that anyone insists on rigid adherence to the traditional threat model, so much as we lack clarity as to 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.

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 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 trade-offs are not tolerable flaws if one seeks to address the question of Internet threats. The rest of this article explores modifications of the traditional threat model.

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.

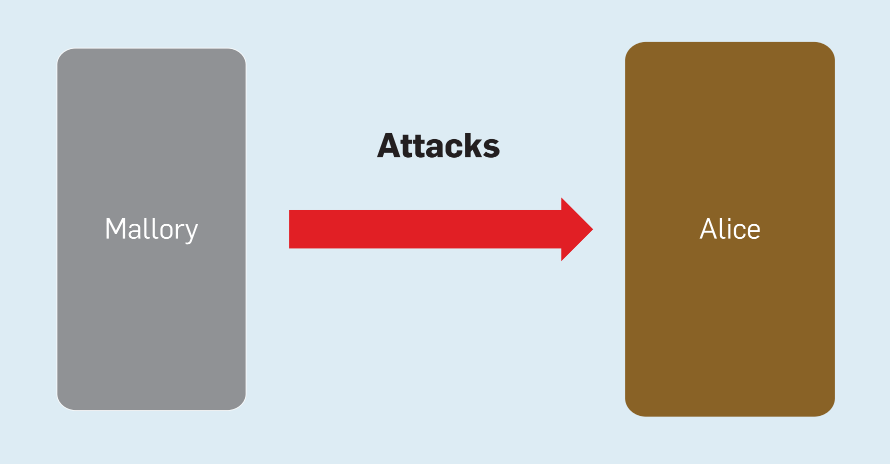
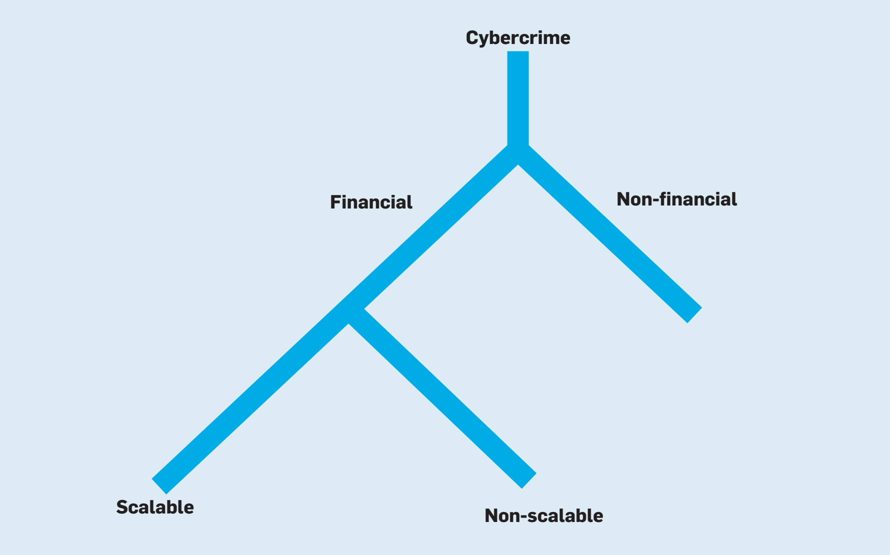


Figure 2. Dividing attacks as financial and nonfinancial; here, financial attacks are further divided into scalable and non-scalable.



Financially Motivated Cybercrime

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 does not distinguish between finite and infinite gain or between zero and non-zero cost. Abstraction like this is useful. It is far more powerful if we can solve the general problem without resorting to specifics. Unfortunately, the attempt breaks down at scale; the binary view of security must be qualified.

When money is the goal, it seems reasonable to assume Mallory is "sufficiently motivated" when the expected gain from an attack exceeds the cost. I 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. I propose a bifurcation of attacks (see Figure 2) into those that are financially motivated and those that are not.

Profit at scale. A main reason for 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: decide who and what to attack, successfully attack, or get access to a resource, and monetize that access.

A particular target is clearly not worthwhile if gain minus cost is not 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:

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

C must include all costs, including that of finding viable victims and of monetizing access to whatever resources Mallory targets. The gain must be averaged across all attacks, not only the successful ones. If either $E\{G\} \rightarrow \infty$ or $C = 0$, then equation (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 small enough to neglect. Against this view I present 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 discussed later, while scripts might achieve access to resources at low cost, the task of monetizing access is generally very difficult. Thus, I argue that not only is attacker cost greater than zero, it is the principal brake on attacker effort.

Attacks that scale. While I have argued that $C > 0$, it is clear that the majority of users are regularly attacked by attacks that involve very low cost per attacked user. I find it useful to segment attacks by how their costs grow. Scalable attacks are one-to-many attacks with 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]

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

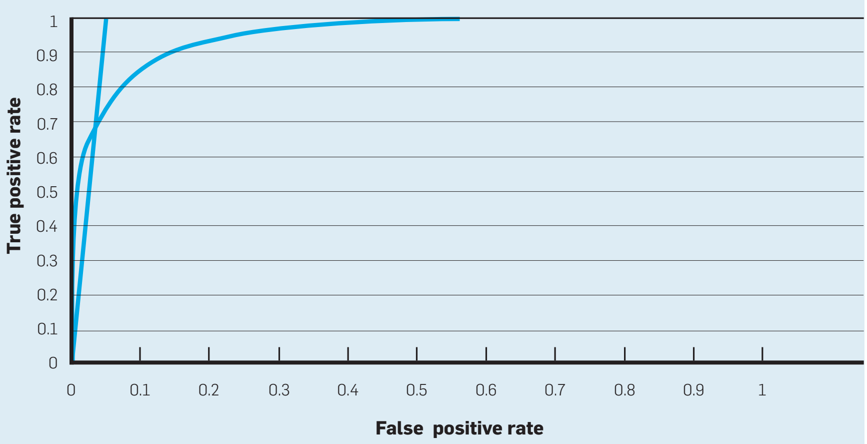
Many of the attacks most commonly seen on the Internet are of this type. Phishing and all attacks for which spam is the spread vector are obvious examples. Viruses and worms that spread wherever they find opportunity are others. Drive-by download attacks (where webpage visitors are attacked through browser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

Constraints on Financially Motivated Attackers

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

Scalable attacks (attack everybody). An alternative to solving the problem of deciding whom to attack is to attack everyone. Scalable attacks have inherent advantages over non-scalable attacks. They reach large masses at very low cost, and techniques can be propagated easily—advantages that come with severe constraints, however. Scalable attacks are highly visible; in reaching millions going unnoticed is

Figure 3. Example ROC curve with line of slope *T/d* = 20. Only operating points to the left of this line satisfy equation (5) and yield profit. 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, less than 1% of viable targets will be attacked.



difficult. Their broadcast nature gives an alert, to both defenders and other would-be attackers. This attracts competition and increases defense efforts.

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 cannot be automated completely or involves per-target effort is thus non-scalable, as this cost violates the constraint defined in equation (2). Physical side-channel attacks (requiring proximity) are out, as getting close to one million users costs a lot more than getting close to one. Labor-intensive social engineering attacks (such as those described by Mitnick[12]) 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 (such as "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 backup authentication questions that involve researching where a user went to high school are non-scalable.

While Internet users see evidence of scalable attacks every day, it is actually a minority of attack types that are scalable.

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, as in equation (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 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 whom 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 to identify targets in a large population, non-scalable attacks are of little use to Mallory.

Assume 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 average cost). Call this $P\{\text{viable}|\text{obs.}\}$. The observables might be address, ZIP code, occupation, and any other factor likely to indicate profitability. Without loss of generality, they can be wrapped into a single one-dimensional sufficient statistic.[15] We assume the cost of gathering the observables is small relative to the cost of the attack. This makes the problem a binary classification,[9] so 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 us now 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's theorem (when d is small):

$$P\{\text{viable}|\text{obs.}\} = \frac{d}{d + \dfrac{P\{\text{obs.}|\text{ non-viable}\}}{P\{\text{obs.}|\text{viable}\}} \cdot (1-d)}$$

$$\approx d \cdot \frac{P\{\text{obs.}|\text{viable}\}}{P\{\text{obs.}|\text{non-viable}\}}. \qquad (3)$$

$P\{\text{viable}|\text{obs.}\}$ is proportional to density; so the difficulty 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 promise a near "sure thing" at one density offer a worse than 1,000-to-1 long shot at another.

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 if, say, an attacker needs one attack in every $1/T$ (such as 1-in-20 and 1-in-100) to be profitable. Then, from equation (3) he must have:

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

This constraint says 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 that are T/d = 500x 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 (when d is small) we get:

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

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 in Figure 3.

Since the slope of the ROC curve is monotonic,[15] as we retreat to the left, t_p/f_p thus increases; equation (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 a smaller and smaller fraction of the true positives, or viable targets, are attacked; for example, for the classifier in Figure 3, when $T = 1/10$ (or Mallory needs one attack in 10 to succeed) and $d = 10–5$ (or one in 100,000 is viable), Mallory requires $t_p/f_p \geq 104$, which happens only for values $t_p < 0.01$, meaning 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] Without a very good classifier (with 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] Even with the benefits of decades of effort and millions of examples of both true and false positives, building a classifier is often extremely difficult. 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. Determining how viable and non-viable can be distinguished requires a large collection of viable targets.

Monetization: Access ≠ dollars. In many forms of non-financial cybercrime 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 is in he is done. A few screenshots, a decorated webpage, or extruded files suffice if the attacker merely wants acknowledgment. However, for financially motivated attacks, things are different. The attacker is not 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 they lead to money. Turning access into money is much more difficult than it looks.

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

The situation with other account types is typically worse. Attempts to monetize access to social-networking passwords generally involve the 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 we already found problematic.

When resources are finite, the question is not whether trade-offs will be made, but how.

Other consumer assets also seem challenging. It may be possible to compromise users' machines by getting them to click on a malicious link. However, even with arbitrary code running on the machine, monetization is far from simple. All passwords on a machine can be harvested, but we have seen that only a minority of stolen bank passwords can be monetized and most nonbank passwords are worthless. The machine can be used to send spam, but the return on spam-based advertising campaigns is low.[10] A botnet responsible for one-third of the world's spam in 2010 apparently earned its owners $2.7 million.[1] A machine can be used to host malicious content. However, as an argument for monetization, this logic is circular, suggesting how yet more machines can be infected, rather than how the original or subsequent machines can be monetized. Scareware, or fake anti-virus software, appears to be one of the better prospects. Successfully compromised boxes can be sold; a pay-per-install market reportedly pays on the order of $100 to $180 per thousand machines in developed markets.[2] Ransomware offers another possible strategy but works best against those who do not practice good backup regimes. For a financially motivated attacker, bank passwords seem to be the best of the consumer-controlled assets, though that best is not very good.

Popular accounts often paint a picture of easy billions to be made from cybercrime, though a growing body of work contradicts this view. Widely circulated estimates of cybercrime losses turn out to be based on bad statistics and 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.[10] In a meta-analysis of available data, Anderson et al.[1] estimated global revenue 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), popular in the days of dial-up-modem access, have resurfaced in mobile-phone malware. It would be wrong to conclude 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 for those who specialize in infrastructure, selling services to those downstream, capture much of the value.

The difficulty of monetization appears to be not clearly understood. The idea that attacks resulting in nonfinancial harm might have been worse is quite common. The journalist Mat Honan of *Wired* magazine, whose digital life was erased but who suffered no direct financial loss, said, "Yet still I was actually quite fortunate. They could have used my email 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 account and an iTunes account and wiped several devices. While exceedingly inconvenient for the victim, anyone attempting to monetize these accomplishments would likely be disappointed.

Discussion

Scalability is not a "nice to have" feature. Today's 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 criminals in 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 fall into only a few limited categories. Non-scalable attacks face a different problem. While the number of viable victims in even a niche opportunity may be large, the difficulty of finding them is related to their relative frequency, not their absolute number. In this case, while the attack itself is non-scalable, Mallory still needs a low-cost way to accurately identify the good prospects in a vast population.

Trade-offs are not optional. When resources are finite, the question is not whether trade-offs will be made but how. For defenders, a main problem with the traditional threat model is it offers no guidance whatsoever as to how it 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.

Most assets escape exploitation not because they are impregnable but because they are not targeted.

We are unlikely to be able to defeat unconstrained attackers, who, according to Pfleeger and Pfleeger,[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 understanding target selection and monetization constraints is necessary if we are to make the unavoidable trade-offs in a systematic way.

Which attacks can be neglected? As before, I concentrate on attacks that are financially motivated, where expected gain is greater than cost. Scalable attacks represent an easy case. Their ability to reach vast populations means no one is unaffected. They leave a large footprint so are not difficult to detect, and there is seldom much mystery as to whether or not an attack is scalable. In the question of trade-offs it is difficult 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 earlier. Everyone needs to defend against them.

Non-scalable attacks present our opportunity; it is here we must look for candidates to ignore. Cybercriminals probably do most damage with attacks they can repeat and for which they can reliably find and monetize targets. I suggest probable harm to the population as a basis for prioritizing attacks. First, attacks where viable and non-viable targets cannot be distinguished pose the 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 small T/d becomes very large, and the fraction of the viable population that is attacked shrinks to nothing, or $t_p \rightarrow 0$. This suggests non-scalable attacks with low densities are smaller threats than those where it is high. Finally, the more difficult an attack is to monetize the smaller the threat it poses.

Examples of attacks with low densities might be physical side-channel attacks that allow an attacker in close proximity to the target to shoulder surf and spy on the output on a screen or printer or the input to a keyboard. The viable target density would be the fraction of all LCD screens, printers, and keyboards whose output (or input) can be successfully attacked and monetized for greater reward than the cost of the attack. It seems safe to say 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 difficult to monetize are low-value assets (such as email messages and social networking accounts); while these occasionally lead to gain, the average value appears quite low.

Analysis of the observability, density, and monetization of attacks will never be perfect. To some degree, judgments must be retroactive. That errors will be made seems unavoidable; however, since we are unable to 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 (such as small businesses in the U.S.) or categories of attacks (such as spear phishing email messages) are clearly being profitably exploited, additional countermeasures are warranted.

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 makes it impossible to satisfy equation (2). Imposing a small charge has been suggested as a means of combatting spam,[3] and it is worth considering whether it might be applicable to other scalable attacks. Address space layout randomization similarly converts scalable attacks to non-scalable. Relatively unexplored is the question of how to make the classification problem even more difficult. 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 more difficult; 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 email messages it reduces return and makes it more difficult to make a profit.[9]

I have repeatedly stressed that an attacker must choose targets, successfully attack, and then monetize the success. The second of these problems has dominated the research effort. However, if the admonition "Think like an attacker" is not empty, we should pay equal attention to how attackers can select targets and monetize resources. I have pointed out the theoretical difficulty of the binary-classification problem represented by target selection. Yet for a profit-seeking attacker the problem is not abstract. It is not enough to hopefully suggest that some ZIP codes, employers, or professions might be indicative of greater viability than others. The attacker needs concrete observable features to estimate viability. If he does not get it right often enough, or does not satisfy equation (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 difficult to justify spending on defense. I reiterate that I have focused on financially motivated attacks. An interesting research question would be which types of target are most at risk of non-financial attacks.

Conclusion

When we ignore attacker constraints, we make things more difficult than they need to be for defenders. This is a luxury we cannot afford. The view of the world that says every target must block every attack is clearly wasteful, and most of us understand 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 trade-offs are not possible, and reality, which says they are necessary, must be resolved. I propose the difficulties of profitably finding targets and monetizing them are underutilized tools in the effort to help users avoid harm.

Acknowledgments

I thank Rainer Böhme, Joe Bonneau, Shuo Chen, Alain Forget, Rob Reeder, Adam Shostack, and the anonymous reviewers for comments that greatly improved this article.

References

1. Anderson, R., Barton, C., Böhme, R., Clayton, R., van Eeten, M. J.G., Levi, M, Moore, T., and Savage, S. Measuring the cost of cybercrime. In *Proceedings of the 11th Annual Workshop on the Economics of Information Security* (Berlin, June 25–26, 2012).
2. Caballero, J., Grier, C., Kreibich, C., and Paxson, V. Measuring pay-per-install: The commoditization of malware distribution. In *Proceedings of the USENIX Security Symposium*. USENIX Association, Berkeley, CA, 2011.
3. Dwork, C. and Naor, M. Pricing via processing or combatting junk mail. In *Proceedings of Crypto 1992*.
4. Elmore, J.G., Barton, M.B., Moceri, V.M., Polk, S., Arena, P.J., and Fletcher, S.W. Ten-year risk of false positive screening mammograms and clinical breast examinations. *New England Journal of Medicine 338*, 16 (1998), 1089–1096.
5. Florêncio, D. and Herley, C. Is everything we know about password-stealing wrong? *IEEE Security & Privacy Magazine* (Nov. 2012).
6. Florêncio, D. and Herley, C. Sex, lies and cyber-crime surveys. In *Proceedings of the 10th Workshop on Economics of Information Security* (Fairfax, VA, June 14–15, 2011).
7. Graff, L., Russell, J., Seashore, J., Tate, J., Elwell, A., Prete, M., Werdmann, M., Maag, R., Krivenko, C., and Radford, M. False-negative and false-positive errors in abdominal pain evaluation failure to diagnose acute appendicitis and unnecessary surgery. *Academic Emergency Medicine 7*, 11 (2000), 1244–1255.
8. Herley, C. The plight of the targeted attacker in a world of scale. In *Proceedings of the Ninth Workshop on the Economics of Information Security* (Boston, June 7–8, 2010).
9. Herley, C. Why do Nigerian scammers say they are from Nigeria? In *Proceedings of the 11th Annual Workshop on the Economics of Information Security* (Berlin, June 25–26, 2012).
10. Kanich, C., Weaver, N., McCoy, D., Halvorson, T., Kreibich, C., Levchenko, K., Paxson, V., Voelker, G.M., and Savage, S. Show me the money: Characterizing spam-advertised revenue. In *Proceedings of the 20th USENIX Security Symposium* (San Francisco, Aug. 8–12). USENIX Association, Berkeley, CA, 2011.
11. Lampson, B. Usable security: How to get it. *Commun. ACM 52*, 11 (Nov. 2009), 25–27.
12. Mitnick, K. and Simon, W.L. *The Art of Deception: Controlling the Human Element of Security*. John Wiley & Sons, Inc., New York, 2003.
13. Pfleeger, C.P. and Pfleeger, S.L. *Security In Computing*. Prentice Hall Professional, 2003.
14. Schneider, F. Blueprint for a science of cybersecurity. *The Next Wave 19*, 2 (2012), 47–57.
15. van Trees, H.L. *Detection, Estimation and Modulation Theory: Part I*. John Wiley & Sons, Inc., New York, 1968.

Cormac Herley (cormac@microsoft.com) is a principal researcher in the Machine Learning Department of Microsoft Research, Redmond, WA.