Where Do All The Attacks Go?

Dinei Florêncio and Cormac Herley
Microsoft Research
One Microsoft Way
Redmond, WA, USA
dinei@microsoft.com, cormac@microsoft.com

ABSTRACT

The fact that a majority of Internet users appear unharmed each year is difficult to reconcile with a weakest-link analysis. We seek to explain this enormous gap between potential and actual harm. The answer, we find, lies in the fact that an Internet attacker, who attacks *en masse,* faces a sum-of-effort rather than a weakest-link defense. Large-scale attacks must be profitable in expectation, not merely in particular scenarios. For example, knowing the dog's name may open an occasional bank account, but the cost of determining one million users' dogs' names is far greater than that information is worth. The strategy that appears simple in isolation leads to bankruptcy in expectation. Many attacks cannot be made profitable, even when many profitable targets exist. We give several examples of insecure practices which should be exploited by a weakest-link attacker but are extremely difficult to turn into profitable attacks.

1. INTRODUCTION: WHY ISN'T EVERYONE HACKED EVERY DAY?

Internet security has a puzzling fact at its core. If security is only as strong as the weakest-link then all who choose weak passwords, re-use credentials across accounts, fail to heed security warnings or neglect patches and updates should be hacked, regularly and repeatedly. Clearly this fails to happen. Two billion people use the Internet; the majority can in no sense be described as secure, and yet they apparently derive more use from it than harm. How can this be? Where do all the attacks go?

We do not have to look far for evidence that things are bad. The range of attacks to which Internet users are subjected is enormous. Attack vectors seldom disappear, and new threats emerge all the time. Brute-forcing, Man-in-the-middle attacks and session hijacking have been with us for some time, but have recently been joined by a host of new threats. Phishing emerged in the last decade. While it has not declined, exploits such as Cross-Site Request Forgery and keylogging Trojans have been added to the list. The previously unknown phenomenon of botnets has mushroomed into prominence in the last five years. In the last few years we have learned that DNS, on which the name structure of the Internet depends, Chip-and-PIN, which handles hundreds of millions of transactions per day, and SSL, which handles encrypted traffic online "are broken." [26, 35].

Against this backdrop, there are approximately 2 billion people using the Internet [5]. Larger services like Facebook, Yahoo! and Hotmail have hundreds of millions of users each. It is not speculation to say that the majority of Internet users ignore the majority of security advice they are offered. In spite of the large and growing set of attacks, numerous studies show that users choose weak passwords [11, 15], ignore certificate error warnings [34], cannot tell phishing sites from legitimate ones [12], are careless about the status of their anti-virus protection and re-use passwords across accounts liberally. A recent report by Webroot [1] found that 90% share password across accounts, 41% share passwords with others, 40% never use special characters in passwords, and 14% have never changed their banking password. Updating software, regarded as a vital security practice, is largely left to chance. As of Sept. 2010, fully 58% of Windows users were still running Windows XP [2], and 22% of Internet Explorer users still use IE6 more than four years after the launch of IE7, and a year and a half after IE8. Trustseer reported in 2009 that 80% of users were running un-patched versions of Flash [3]. Users are not alone in this negligence: Rescorla reports that even among system administrators fewer than 40% had installed a long-available patch against the Slapper worm [33].

Yet, if things are so bad, how come they're so good? It is not speculation to say that the majority of users are not harmed every day. Estimates place the number of users who have accounts hijacked each year at below 5% [16, 14]. So, 95% or more of users suffer no harm from account hijacking each year. Thus, most users fall well short of the effort required to "be secure" and yet they mostly escape harm. For example, the majority of the 90% from the Webroot survey who re-use passwords

across accounts almost certainly escape harm. Equally, while Chip-and-PIN may be broken, it is involved in several hundred million transactions per day with apparently manageable levels of fraud. The great potential for harm of the Slapper worm was never fulfilled. Close to 50% of the DNS infrastructure remained unpatched at time of Kaminsky's disclosure, and yet, for all intents and purposes, nothing happened.

So, where do all the attacks go? In this paper we seek to explain this enormous gap between potential and actual harm. A weakest-link analysis seems unable to offer an explanation. The answer, we suggest, lies in a shortcoming of common threat models. The model where a single user Alice faces an attacker Charles fails to capture the anonymous and broadcast nature of web attacks. Indeed, it is numerically impossible: two billion users cannot possibly each have an attacker who identifies and exploits their weakest-link. Instead, we use a cloud threat model where a population of users is attacked by a population of attackers. Our main finding is that a crowd of users presents a sum-of-effort rather than a weakest-link defense. Many attacks, while they succeed in particular scenarios, are not profitable when averaged over a large population. This is true even when many profitable targets exist and explains why so many attacks types end up causing so little actually observed harm. Thus, how common a security strategy is, matters at least as much as how weak it is. Even truly weak strategies go unpunished so long as the cost of the failures exceeds the gain from the successes. Why is this question important? If, as appears to be the case, a majority are insecure and yet unharmed it is important to understand why. These users are avoiding harm at far lower cost than is usually assumed to be necessary.

2. A THREAT MODEL THAT SCALES

System-centric threat models often describe the technical capabilities of an attacker. A defender Alice is pitted against an attacker Charles, who can attack in any manner consistent with the threat model. Generally Alice's security is regarded as being only as good as the weakest-link.

There are several things wrong with this threat model. It makes no reference of the value of the resource to Alice, or to Charles. It makes no reference to the cost of defence to Alice, or of the attack to Charles. It makes no reference to the fact that Charles is generally uncertain about the value of the asset and the extent of the defence (*i.e.*, he doesn't know whether benefit exceeds cost until he attacks successfully). It makes no provision for the possibility that exogenous events save Alice, even when her own defence fails (*e.g.*, her bank catches fraudulent transfers). It ignores the fact that Charles must compete against other attackers (we showed how this drives down returns in previous work [22]). It ignores scale: assuming that Internet users greatly outnumber attackers it is simply numerically impossible for every user to have an attacker who identifies and exploits her weakest-link. Some high-value users may face this threat model, but it is not possible that all do. Some or all of these shortcomings have been addressed by others; see the Related Work section for details. It is however one last failing that we are primarily interested in addressing. This model, where weakest-links are ruthlessly exploited, is unable to explain the reality we observe: 20% use a significant date or pet's name as password, yet 20% are not victimized. It is this inability to explain observations that we seek to address.

2.1 An Internet threat model

In our threat model a population of Internet users are attacked by a population of hackers. We call the Internet users Alice(i) for $i = 0, 1, \cdots, N^{'} - 1$ and the attackers Charles(j) for $j = 0, 1, \cdots, M - 1$. Clearly $N^{'} \gg M$: Internet users outnumber attackers. Each Alice(i) is subjected to attack by many of the attackers. Each attacker goes after as many Internet users as he can reach. Cost is the main reason for this approach: it costs little more to attack millions than it does to attack thousands. The attackers' goal is purely financial. None of the Alice(i)'s are personally known to any of the Charles(j)'s. Thus, revenge, jealousy, curiosity and emotion play no role. The situation is depicted in Figure 1. This threat model captures the large-scale broadcast attacks so familiar to Internet users: phishing, malware-bearing spam, for example. These attacks are similar to the parallel attacks that Schechter and Smith mention [39] and the scalable attacks that Herley studies [21] and the distributed attack network that Fulz and Grossklags study [29]. For more details on related work see Section 6.

Our threat model differs from others in several respects. First, we focus on end-users of the Internet. Thus we examine the consumer rather than the enterprize space. This is significant for a number of reasons. Consumers generally have less information and are less protected. They must decide for themselves a whole range of issues that affect their security from passwords to anti-virus to software updates. But they do so largely in ignorance. They do not have security professionals who monitor their network searching for problems and anomalies. They do not have well developed expectations as to where their weakest-links lie. Even after Alice(i) has an account hijacked or money stolen she has very little ability to carry out forensic examination and determine what happened. Second, rather than having an individual defender and individual attacker pitted against each other, we have a population of $N^{'}$ users facing M attackers. Attackers must strive, not

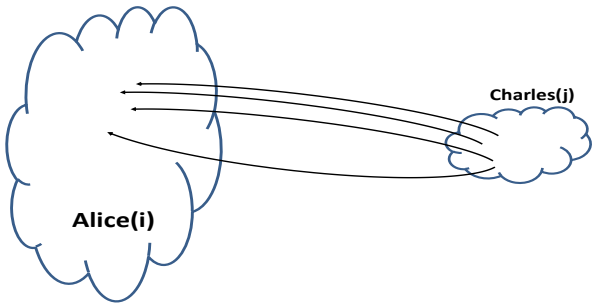

Figure 1: Threat Model: a population of Internet users Alice(i) are attacked by a population of hackers Charles(j). Each user, Alice(i) receives attacks from numerous different attackers, each hacker Charles(j) attacks many different users. If Charles(j) successfully obtains access to Alice(i)'s account he then attempts to monetize the asset.

merely to attack users, but also to compete with each other. While $N^{'}$ is large it is finite. Charles(j) faces the prospect that the most easily-attacked users will be victimized by several attackers. There is a chance that Charles(j) successfully attacks Alice(i) only to find that Charles($j-1$), Charles($j-2$) and Charles($j-3$) have already been there. As all Internet users know: as far as spam, phishing, *etc*, are concerned there are no un-contacted populations. Third, the attacks we study happen in a competitive economic landscape. An attack is not merely a technical exploit but a business proposition. If it succeeds (and makes a profit) it is repeated over and over (and copied by others). If it fails (does not make a profit) it is abandoned and the energy is spent elsewhere. Fourth, attackers are playing a "numbers game": they seek victims in the population rather than targeting individuals. For example, if Charles(j) targets Paypal accounts, he isn't seeking particular accounts but rather any accounts that he happens to compromise. Charles(j) doesn't know the value, or security investment of any particular Internet user in advance. He discovers this only by attacking.

We freely admit that this threat model has some obvious short-comings. It excludes cases where the attacker and defender are known to each other, or where non-monetary motives are involved. It does not cover cases of attackers motivated by emotion, curiosity, revenge or the desire for fame or notoriety. It does not cover the case of Advanced Persistent Threats. It does not cover the case where the attacker is targeting Alice(i) alone or values her assets beyond their economic value. While restrictive, our model of an unknown, financially motivated attacker does cover a significant fraction of what most users are concerned with.

2.2 Expected gain and expected loss

Our Internet user Alice(i) has assets that she must protect. For any particular asset there are many possible attacks, call them attack(0), attack(1), $\cdots$, attack($Q-1$). For example, keylogging, phishing, brute-forcing are all methods of attacking an Internet account. An attacker can choose whichever gives the best return for his effort.

We model Alice(i)$'$s expected loss as follows. The effort that Alice(i) devotes to defending against attack(k) is $e_i(k)$. $Pr\{e_i(k)\}$ is the probability that she succumbs to this attack (if attacked) at this level of effort. We assume that $Pr\{e_i(k)\}$ is a monotonically decreasing function of $e_i(k)$. This merely means that the greater effort Alice(i) spends on attack(k), the lower her probability of succumbing. L_i is the loss that Alice(i) endures when she succumbs to any attack, independent of the attack type. For example, it doesn't matter whether her password was stolen by keylogging or brute-force. In addition, to allow for external fraud checks, there is some chance that, even though she succumbs to attack, Alice(i) suffers no loss because she is saved because of exogenous events. For example, her bank password falls to Charles(j) but her bank detects the fraud and saves her from harm. Here, we use $Pr\{\text{SP}\}$ to denote the probability that her Service Provider saves Alice(i) from harm. Alice(i)'s loss then is the probability that exogenous events do not save her, times the probability that she succumbs to any of the attacks, times her loss, plus the sum of what she spends defending against all attacks:

$$(1 - Pr\{\text{SP}\}) \cdot \left(1 - \prod_{k=0}^{Q-1} (1 - Pr\{e_i(k)\})\right) L_i + \sum_{k=0}^{Q-1} e_i(k). \quad (1)$$

The goal of Alice(i) is to minimize her expected loss under the range of attacks that she sees.

On the other side of the fence what is the expected gain for an attacker Charles(j)? We denote G_i as his gain (if successful) from Alice(i), and $C_j(N,k)$ as the cost to Charles(j) of reaching N users with attack(k). The expected gain of the attacker Charles(j) is the probability that exogenous events do not stop his fraud, times the sum of the probable gain over all attacked users, minus the total cost of the attack:

$$U_j(k) = (1 - Pr\{\text{SP}\}) \cdot \left(\sum_i Pr\{e_i(k)\} G_i\right) - C_j(N,k). \quad (2)$$

The summation in (2) is over as many users, N, as Charles(j) attacks. We don't assume that all $N^{'}$ Internet users are attacked, however we assume that the

number is large enough for statistical arguments to apply. This accords with our threat model: many Internet attacks have costs that grow far slower than linearly with the number of users attacked, so it makes sense to attack as many users as possible. The spam campaign documented by Kanich *et al.* [27], for example, attacked 350 million users. So assuming that Charles(j) attacks at least thousands is not overly restrictive. It also bears mentioning that many attacks might have a fixed cost that is almost independent of the number of users attacked. Charles(j) might be able to spam 350 million users for $100, but he can't reach 3.5 million for $1.

G_i is the gain that Charles(j) extracts from Alice(i). Now, Charles(j)'s gain, G_i, is not necessarily the same as Alice(i)'s loss, L_i. There are several reasons for this. We assume that the asset is rivalrous [30], which means that enjoyment of it by one party reduces enjoyment of it by another. Thus

$$G_i \leq L_i,$$

so that Charles(j) can at most gain whatever Alice(i) loses. It is possible that Charles(j) is not alone in successfully attacking Alice(i), so that he shares the loss that Alice(i) suffers with several others; *i.e.*, $G_i \approx L_i/m$ for some number of attackers m. We explore this possibility in Section 4.3.

If the asset is non-rivalrous other possibilities exist. First, Charles(j) might benefit without harming Alice(i): *e.g.*, if he uses Alice(i)'s machine simply to send spam and conceal his IP address he might derive significant value while Alice(i) would not suffer directly. Thus, $G_i \gg L_i$. An intriguing possibility, where $L_i < 0$, and Alice(i) makes a "pact with the devil" and benefits from the attack is explored by Bond and Danezis [9]. Finally, it is possible that $G_i \ll L_i$; this might be the case of vandalism. For example, if instead of attempting to monetize the asset Charles(j) set out to destroy it. We won't treat either of these cases further and instead concentrate on the rivalrous case.

3. THE INTERNET ATTACKER FACES A SUM-OF-EFFORTS DEFENSE

Security is often described as a weakest-link game [28, 32], where security depends on the most easily breached part of a defence. This has colored much thinking in the space. It is hard however, to square this with the claim that 20% of users choose a pet's name or significant date as password, and the fact that password re-use across accounts is almost universal [15].

The weakest-link analysis makes perfect sense where a single attacker faces a single defender [28, 32]. Since the game is zero-sum (or negative sum) with only two players the threat that is most profitable for the attacker is the one that is most costly for the defender. However, for the Internet attack model that we are using, where a crowd of users face a crowd of attackers, this is no longer the case. The threat that is most profitable for the attacker need not be any individual user's weakest-link. Further, an individual user's weakest-link need not be exploited by the most profitable attack for any attacker. In fact, as we know show, the simple change in threat model changes the defense that an Internet attacker confronts from a weakest-link defense into a sum-of-efforts one. For example, the fact that users who choose their birthdate as password avoid harm is puzzling in a weakest-link analysis but makes prefect sense in our threat model.

Elements of weakest-link, sum-of-effort and best-shot games are all present in the formulation above, and their roles are crucial as we show now. An excellent analysis of these three games in a security setting is given by Varian [19]. An analysis of how these games differ in a protection and insurance environment is performed by Grossklags *et al.* [18] who also introduce the concept of a weakest-target game.

3.1 Attack Selection

Each attacker Charles(j) chooses the attack that maximizes his expected gain. That is, ranging over all attack(k), he selects $\max_k U_j(k)$. Not all attackers may have the same cost structure, so what is the best attack for one, may not be so for another. For example, for Charles(j) and Charles($j+1$) the best attack might be attack(k), while for Charles($j+2$) it might be attack(k'). This explains the co-existence of several attacks on the same asset class. For example, many different attacks on user credentials co-exist; this suggests that there is no single attack(k) which maximizes the expected gain for all attackers. However, it is likely that some attacks give the best return to a wide set of attackers, while some are best for almost none. It is also likely that this changes with time.

3.2 Sum-of-efforts

Examining (2) we see that for greater than zero gain Charles(j) requires that his return exceeds his costs:

$$(1 - Pr\{\text{SP}\}) \cdot \left(\sum_i Pr\{e_i(k)\} G_i \right) > C_j(N, k).$$

Recall that $Pr\{e_i(k)\}$ is a decreasing function of user effort $e_i(k)$. The left-hand side is related to the sum-of-efforts of all attacked users, weighted by the gains. The greater the total effort of the user population the lower the return. Thus, the expected gain from any attack is a *sum-of-effort* game [19]. An attack can be unprofitable (*i.e.*, $U_j(k) < 0$) if the sum-of-effort of users is great enough, even though individual users represent good targets. We examine this further in Section 4.1. The formulation of (2) is not precisely sum-of-effort. Increasing effort by those who are above the threshold to

escape harm does nothing to reduce the return. Thus it is effectively a non-linear sum-of-efforts defense.

A sum-of-efforts defense is known to be far more effective than weakest-link. The well-known free-rider effect [23] ensures that many users escape harm, even at low levels of effort. This will play an important role in the remainder of the paper.

3.3 Best-shot

Detection of fraud by the service provider is a *best-shot* game. That is, if any of a series defences catches the fraud, then Charles(j) fails. For example, a bank may have a series of checks in place to detect fraud. Accounts that have little history of outbound transfer, logins from geographies outside the user's pattern, transfers to a stranger's account may all alert suspicion. The success of credit card fraud detection illustrates that $Pr\{SP\}$ can be quite high based purely on customer usage patterns. If in the process of attempting to drain the account Charles(j) triggers any of them then his gain is zero. In fact Charles(j) faces a sum-of-effort defense, cascaded with a best-shot defense. That is, he must succeed first against a sum-of-effort defense (to successfully compromise enough users). Following this, he must succeed against a best-shot defense (to successfully evade the fraud detection measures of the service provider).

3.4 Contrast between Internet attacker and individual attacker

The difference between a sum-of-effort and weakest-link defenses is so great that it's worth reiterating how it comes about in our threat model. Our Internet attacker faces a crowd of users. He selects attack(k) that maximizes:

$$(1 - Pr\{\text{SP}\}) \cdot \left(\sum_{i=0}^{N-1} Pr\{e_i(k)\} G_i \right) - C_j(N,k).$$

By contrast the individual attacker is after a particular user, Alice(i_0), rather than a crowd. He thus selects attack(k) that maximizes:

$$(1 - Pr\{\text{SP}\}) \cdot Pr\{e_{i_0}(k)\} G_{i_0} - C_j(1,k).$$

This is clearly maximized by the attack for which $Pr\{e_{i_0}(k)\}/C_j(1,k)$ is highest. This is Alice(i_0)'s weakest-link: the highest probability of success/cost ratio. Facing such an attacker Alice(i_0) can indeed afford to neglect no defense. Even slight weaknesses can be exploited. Why then doesn't our attacker target each user in turn? The answer, of course, is his cost structure. Our Internet attacker gets to attack N users with attack(k) at a cost of $C_j(N,k)$. However he cannot afford to target users individually $C_j(1,k) \gg C_j(N,k)/N$. The circumstances that ensure our Internet attacker faces a sum-of-effort rather than weakest-link defense are intrinsic to his *modus operandi.* This is a key point of difference between our model and that produced by the weakest-target game [18]. As the name suggests, those who have invested least succumb in a weakest-target game. However, in our model even those who have invested little or no effort escape harm, so long as there aren't enough such users to make the overall attack profitable in expectation.

4. WHY DO ATTACKS FAIL?

We now turn to the question of why so many exploits and vulnerabilities fail to translate into harm experienced by users. One obvious reason why an attack may never inflict harm is that it has negative return, that is the expected gain is lower than the expected cost. While we often get the impression that cyber-criminals get money "for free" clearly they have costs, just as any legitimate business does. Looking at (2) we can determine several ways that expected gain can be negative. This requires:

$$(1 - Pr\{\text{SP}\}) \cdot \left(\sum_{i} Pr\{e_i(k)\} G_i \right) < C_j(N,k). \quad (3)$$

We now go through several of the possibilities that can satisfy this condition.

4.1 Average success rate is too low

The left-hand side of (3) is the expected return to Charles(j) of attack(k). The sum is an average of the gains to be had, G_i, weighted by the success likelihoods $Pr\{e_i(k)\}$. Each user makes some effort against attack(k); the greater the effort Alice(i) makes the smaller the probability that she succumbs to attack(k). Since $Pr\{e_i(k)\}$ is a monotonically decreasing function of $e_i(k)$, the greater the total effort of the user population the lower the expected return for Charles(j). Thus, if the average effort increases, average success decreases and expected gain decreases. If average success decreases enough (*i.e.*, $1/N \cdot \sum_i Pr\{e_i(k)\} \to 0$), then attack($k$) is unprofitable. It is not necessary that every user increase effort merely that enough of them do.

This leads to a simple explanation of why some attacks fail to happen: the average success rate is low enough to make it uneconomic. Since $Pr\{e_i(k)\}$ is a monotonically decreasing function of $e_i(k)$ this means that average effort is too high. This might seem a far-fetched possibility given what we know of Internet user behavior. However, some attacks require only effort that we already know most users make. For example, if the attack is to password-guess using the top ten passwords from the Rockyou dataset we know that these passwords account for about 2% of accounts. Thus 98% of users have made enough effort to resist this attack.

Consider an attack which is easy to defend against (*i.e.*, for a small effort $e_i(k) > \epsilon$ then $Pr\{e_i(k)\} \approx 0$).

The vast majority of users invest the effort to evade the attack, but a very small number do not. This attack works very well against a tiny fraction of people. However, Charles(j) can determine whether it works only by trying it (*i.e.*, investing the cost $C_j(N,k)$). If the attack works on too small a fraction of the population the attack is uneconomic.

Observe if the attack becomes uneconomic, that users who do not invest enough (*i.e.*, $e_i(k) < \epsilon$ and hence $Pr\{e_i(k)\} \approx 1$) nonetheless escape this attack. Since the average effort is enough to keep the average success low and make the attack unprofitable everyone escapes harm, both those who have invested adequately and those who have not. That sum-of-effort games allow for a free-rider effect is well known [23].

Thus, one major reason that some attacks fail to manifest themselves is that the attack succeeds only on a tiny fraction of the population. A very small set of people use their dog's name as password. A similarly small fraction also use the name of their cat, significant other, child, parent, favorite team, movie star or singer, or use a birthday or anniversary date. A small percent of people who make one of these choices have the name or date (*i.e.*, the dog's name or birthday) discoverable in an automated way. Thus the success of any of these attacks is a small percent of a small percent. If Alice(i) follows one of these strategies it might seem that Charles(j) gets G_i for a small amount of effort. That is, if

$$(1 - Pr\{\text{SP}\}) \cdot Pr\{e_i(k)\}G_i > C_j(N,k)/N \quad (4)$$

doesn't Charles(j) make a profit? This is not so. It is not the case that Charles(j) is attacking Alice(i) at a cost of $C_j(N,k)/N$, but rather that he is attacking a population of N users at a cost of $C_j(N,k)$. If the average gain across the attacked users is not positive then his attack fails. To pick and choose the best targets requires that Charles(j) knows in advance which users have invested least.

4.2 Average value is too low

For attack(k) to be unprofitable we saw that (3) had to hold. In addition to the possibility that the success rate is too low, there is the chance that the average value extracted G_i is too low. That is, Charles(j)'s expected return gives him the probability-weighted average of the gain expected from each user (*i.e.*, the summation in the left-hand side of (3)). If the average value of G_is is too low then the attack again fails to be profitable. Much the same dynamic is at work as in the previous section. The attacker gets the average return for the average cost: the fact that individual users represent profitable targets is of no use if they cannot be identified.

Which attacks fall into this class? A common class is those that are predicated on leveraging a low value asset into a high return. We explore these in Section 5.2.

4.3 Attackers collide too often

An important aspect of our threat model is that attackers live in an environment where they compete with each other. While the pool of Internet users is large, it is finite, and attackers compete for a common asset pool. In Section 2.2 we introduced the possibility that attackers collide in pursuing the same asset. In a finite world collisions are inevitable. Indeed Enright *et al.*[13] mention the case of security research teams colliding in their study of the same botnet, raising the question of how either can be sure whether they are measuring the activity of the botmasters or of other researchers! More concretely, Sariou *et al.* [37] find that many malware-infected machines have multiple infections. For example, 71.8% of clients infected with eZula had at least one other infection. Are collisions infrequent enough to be discounted or do they meaningfully affect the analysis? If m attackers collide in successfully attacking Alice(i) then the expected gain must be divided m ways: $G_i = L_i/m$. We now examine how this can happen.

4.3.1 Outcome is deterministic

Consider the case where $Pr\{e_i(k)\}$ is binary, *i.e.*, has value either zero or one depending on the user's effort:

$$Pr\{e_i(k)\} = \begin{cases} 1 & e_i(k) < \epsilon \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Here, any effort greater than ϵ gives perfect immunity to the attack, while any effort below ϵ ensures that Alice(i) succumbs if she is attacked. To be concrete, suppose that the attack is brute-forcing passwords using a list of ten common passwords (*e.g.*, the ten most common passwords from the RockYou [24] dataset). Any user who has chosen one of those ten passwords (*e.g.*, "abcdefg") always succumbs to this attack, while all others escape unscathed. Now, if Charles(j) attempts this attack he ends up successfully hijacking the accounts of all users who chose these ten passwords. However, he is not alone. Every attacker who follows this strategy enjoys exactly the same success: they also enter each of these accounts. Thus, if m attackers follow this strategy we should have $G_i \approx L_i/m$.

A deterministic outcome is the limiting case of something that is more generally true. When attackers collide the expected return is reduced. Thus an estimate of the likelihood of collision is necessary in evaluating the attacker's expected gain. If there are m attackers and each enjoys an independent probability $Pr\{e_i(k)\}$ of compromising Alice(i) in any given attack then the expected number who succeed is $mPr\{e_i(k)\}$. Thus, for any attack where $Pr\{e_i(k)\} > 1/m$ the attacker Charles(j) must expect to share L_i with others. This is quite counter-intuitive, as it implies that victims who have a *high* probability of succumbing to attack(k) do

not increase in value to Charles(j). As $Pr\{e_i(k)\}$ doubles, so does the number of attackers with whom the asset must be shared. The worst victims for Charles(j) are those who deterministically succumb. He ends up burgling a house with m other burglars, or looting a store whose shelves are already bare. This is self-limiting. Unless Alice(i) changes her behavior she ends up, not with one, but with hundreds of attackers in her account.

It is natural to wonder what happens if the first successful attacker shuts the other $m-1$ out. For example, he might change the password, patch the machine or perform other actions to make sure that others do not get access to the asset in the same manner he did. This makes no difference to the expected return: whether the entire asset goes to the first successful attacker or it is shared among them the average return is unchanged.

4.4 Attack is too expensive relative to alternatives

A further reason that attack(k) can fail to ever be observed is that

$$\text{For some } k^{'}: \quad U_j(k) < U_j(k^{'}) \forall j.$$

That is, there's an attack that's better, having either higher expected gain or lower cost.

Consider the example of a realtime MITM attack on a banking session. This threat can take the form of session hijacking, in which the attacker piggy-backs on the legitimate session, or credential replay, in which the attacker sends a time-varying credential within the time window. In either case the attacker must lie in wait until the user actually authenticates. In both cases the attacker only has a single login session to exploit the account. Clearly this attack, which has a time restriction, has greater cost than one that does not. Since the attacker must be ready to exploit the account whenever the opportunity arises there is a constraint on his time. Since all value must be extracted in one session there is no possibility of selling the account for exploitation by others. If the account has a limit on the maximum daily transfer, then this is the maximum that can be transferred out rather than the entire account balance. For all of these reasons, the cost is greater and the gain lower than an attack that involves gathering the password of an account. A password can be used at will, can be sold on, and can be used to login multiple times if necessary to drain the account. Thus if we consider two accounts, one protected by passwords, and one by a one-time password token, there is little reason to attack the latter unless the expected gain from the better protected account is higher. Financial institutions such as Paypal, which make such tokens available to their users fall into this category: while MITM attacks are possible there is little reason to mount them when a less expensive attack on comparably valuable assets exists.

Murdoch *et al.* [35] recently reported an elegant attack on the Chip and PIN protocol used by many European credit card issuers. The attack confuses a point of sale terminal into believing that it has received the correct PIN, even though this is unknown to the attacker. However, since US issued cards are accepted in Europe there is no need to mount this attack. Why assault the PIN when there is no shortage of equivalently valuable targets that do not have the protection can be attacked?

4.5 Exogenous Fraud Detection is Too High

A final factor that decreases Charles(j)'s expected utility is the probability that exogenous events save Alice(i). That is, if $Pr\{SP\} \approx 1$ in (2) then it is exceedingly difficult for Charles(j) to make a profit, irrespective of how Alice(i) behaves. For example, suppose Alice(i)'s bank detects and halts most attempted fraudulent activity. In this case, the true protection is not the effort $e_i(k)$ that Alice(i) expends defending against attack(k), but the back-end protections that the bank has in place. It is difficult for Charles(j) to cover his costs if this is so.

5. WHERE DO ALL THE ATTACKS GO?

We now examine some of the attacks which do not appear to succeed as often as a weakest-link analysis would suggest. Our goal is not to suggest that any of the practices described are advisable. We merely seek to close the gap between what analysis suggests should be happening, and what observation says is actually the case.

5.1 Choosing your dog's name as password

A dismissive description of a typically bad practice is "using your dog's name as password." Indeed Webroot found 20% of users had used a pet's name or significant date as password [1]. Similar strategies involve choosing one's favorite sport's team, actor or cartoon character as password. But this raises the obvious question: if the practice is really so bad, how do 20% of people get away with it? While this is certainly inadvisable we suggest that profiting from this is a lot harder than it looks. While, for some people this may be their weakest-link, enough people do not follow the practice to ensure that the sum-of-effort that an attacker trying to exploit it faces is enough to ensure that it is unprofitable.

Consider a user who has $100 in a bank account protected with her dog's name as password. This might look like an easy $100 for an attacker, but this commits the error of assuming that the attacker can pick the best targets. We saw in Sections 4.1 and 4.2 that an attacker needs that the attack be profitable in expectation. So (3) must hold, not merely (4). Suppose that 1% of users choose their dog's name as banking password. Further suppose that 1% of users have their dog's name

discoverable automatically (*e.g.*, by running a crawling script at a social networking site), and 1% have their bank username discoverable automatically. This means that, in expectation, an attacker can get into one bank account for every million users he attacks. However, (as we saw in Section 4.3) if the attack is this simple (and the outcome deterministic), it will be attempted by many. Suppose that $m = 100$ attackers follow this strategy. Since the outcome is deterministic all of them succeed in the same accounts and fail in the same accounts, so the expected gain drops by another factor of 100. Thus, our attacker sees his average return drop by eight orders of magnitude from the easy money proposition that we began with. We can insulate ourselves from the error of survivor paradox by asking how an attack scales. For example, a dog's name as bank password seems like a sure thing for some attacker. Instead, we might ask how much it would cost to determine the dog's names of one million banking customers and how much that information would be worth.

5.2 Leveraging a low-value account into a high one

It is sometimes claimed that attackers who gain access to a low value email account can use this to get to banking information, or reset banking passwords *etc.* This may indeed be the case, and this approach probably succeeds some of the time. Again, however, a lot more people appear to use a low-value email as their bank contact than have their accounts emptied every year. In question is not whether this account escalation attack ever succeeds (*i.e.*, does (4) hold for at least one user?) but is it profitable on average (*i.e.*, does (3) hold?).

Suppose Charles(j) gains access to n webmail accounts. Some of these are used as the email account of record for banking sites. Some of those will have the bank username included (many banks exclude username in all email communications). Some of those banking sites will mail a password reset link to an email account (though often only after successfully answering secret questions). For each webmail account if all of these conditions are met Charles(j) gains access to a bank account, otherwise he simply gets to read a stranger's email. Thus, a percent of a percent of a percent of webmail accounts will have high value, while the rest are close to worthless. Profit for Charles(j) results only if the value of the percent of a percent of a percent of n webmail accounts that lead to banking information is greater than the cost of acquiring all n webmail accounts.

Just as in Section 4.1 the attacker needs (3) to hold not just (4). That individual users are profitable targets is not in doubt, however Charles(j) attacks his victims in bulk and needs the average gain to be positive. To pick and choose the best targets requires that Charles(j) is omniscient and knows in advance which users have greatest extractable value. It might seem that Charles(j) can boost his return by targeting those with high net-worth. However, high networth and extractable value are not necessarily correlated [21]. Targeting Bill Gates or Warren Buffet is not a sure path to increasing expected gain. In addition, as we saw in Section 3.4, targeting small segments violates Charles(j)'s cost model. He attacks a massive number of users for $C_j(N, k)$, but achieves very little reduction in cost by scaling down.

Again, just as in Section 4.1, sum-of-effort defense implies that there is a free-rider advantage. The average value that can be extracted from an email account is very low. Some email accounts allow access to far more valuable assets and thus represent profitable targets. However, determining which are profitable and which are not cannot be done without mounting (and incurring the cost of) the full attack. If the whole attack becomes unprofitable, then users who have high value escape along with those who have low. Those who have invested least escape, thanks to those who have invested more.

5.3 Domino effect of password re-use

Another frequent claim is that attackers stealing the credentials of one account will exploit the well-known habit of users to re-use passwords across accounts. The thinking is that armed with, for example, a `facebook` password an attacker may be able to gain access to the Wells Fargo account of the user. "One weak spot is all it takes to open secured digital doors and online accounts causing untold damage and consequences" write Ives *et al.* of this possibility [8]. Again, however, we are left with the puzzle that this appears to happen a great deal less than it might. We know that the practice of re-using passwords across accounts is almost universal [1, 15]. If the practice is so common, and so bad why is there not greater evidence of harm?

The answer, we suggest, again, lies in the confusion between having an attack that occasionally works and one that can be made economic at scale. Some `facebook` passwords are doubtless used also for banking. However, determining the correct bank and the username is not straightforward. First, to be successful at scale, determination of the username must be automated: it is clearly impractical for Charles(j) to wade through a thousand compromised `facebook` accounts seeking hints as to the username. This is especially so since he doesn't know that the `facebook` and bank password are the same until he successfully logs in. Thus, the entire process must be automated. Hence, Charles(j) needs not merely that the passwords be the same, but that the bank username either be the same, or be eas-

ily determined in an automated way from the `facebook` account information. If 1% of users satisfy the first criterion and 1% the second then out of a thousand compromised `facebook` accounts Charles(j) has only a 1 in 10 chance of gaining access to a single bank account.

5.4 Fraud Detection

While it is unlikely that $Pr\{SP\} = 1$ in many domains it appears to be high. Persistent reports that credentials sell for fractions of a penny on the dollar [25] indicate that cashing out is hard. The fact that, at least in the US, consumers are protected from the financial consequences of fraudulent transfers and credit-card transactions suggests that banks have considerable ability to detect fraud (*i.e.*, $Pr\{SP\} \approx 1$) even when all else fails.

In fact, since this term applies to all attacks, improving $Pr\{SP\}$ may be a better investment for a bank than any other. This protects all users, whether they are diligent or not. Indeed, highly successful fraud detection assuming that Alice(i) will become compromised may give better return on investment than new technologies that help Alice(i) avoid compromise.

5.5 Diversity is more important than strength

In Section 4.1 we saw that even very poor security practices can go unpunished. If the fraction of users who succumb to a certain attack is too small then the entire attack is unprofitable. When this happens those who would succumb to the attack get a free ride. Those who choose their dog's name as password escape harm simply because not enough people do so to make the attack profitable. Equally, however many other poor security practices go unexploited because of the uneconomic nature of the attack when scaled up to the whole population. This leads to the interesting conclusion that a great many users can have poor security practices that go unexploited so long as a small enough minority follows the same practice. The use of the names of pets, friends, significant others and teams and birthdays as passwords are all bad practices, but each of them is probably rare enough (and hard enough to exploit in an automated way) to make attacking any of them unprofitable. The importance of diversity in computing ecosystems has been recognized since a paper on the subject by Geer *et al.*[17].

Let's look at the implications of the free-rider effect caused by the sum-of-effort nature of the expected return from an attack. If brute-forcing and password guessing is a problem, suppose that $N-1$ of our Internet users finally decide to choose strong passwords for all of their accounts. One user, Alice(i_0), takes a pass and continues her practice of using "abcdefg" as password everywhere. Through no action on her part Alice(i_0)'s risk of harm from brute-forcing decreased dramatically. Brute-forcing is now an infeasible attack for most users and the expected return plummets. In fact, two things determine whether Alice(i_0) succumbs to attack(k). The first is Alice(i_0)'s own effort: the higher $e_{i_0}(k)$ the lower the probability $Pr\{e_{i_0}(k)\}$ that she succumbs *if attacked*. The second is whether she is attacked at all. That is, if attack(k) isn't profitable for any Charles(j) then the attack is never seen at all. One way this can happen is if all other users invest a lot more than Alice(i_0). She gets away with being sloppy, so long as enough users make the effort to make the attack unprofitable. Similarly, all users can be sloppy, so long as they are sloppy in different ways. Schechter *et al.* [36] similarly argue that it is popularity, rather than strength, of passwords that represents a vulnerability.

5.6 Effort Allocation is hard

Just as an attack can be unprofitable for Charles(j), effort can be unprofitable for Alice(i). From (1) the totality of Alice(i)'s effort is profitable only if:

$$(1-Pr\{\text{SP}\})\cdot\left(1-\prod_k(1-Pr\{e_i(k)\})\right)\cdot L_i > \sum_k e_i(k).$$

If this does not hold then Alice(i) is spending more on effort to avoid attacks than her expected loss. Further, her investment in effort against any particular attack(k) is profitable only if

$$(1-Pr\{\text{SP}\})\cdot Pr\{e_i(k)\}\cdot L_i > e_i(k).$$

When this occurs Alice(i) is rational to ignore the effort and run the risk of the harm. This is exactly the rational rejection of security effort against attack(k) described by Herley [20].

Since, Alice(i) does not know what her weakest-link is, effort allocation is extremely hard. If she defends against all attacks she is wasting a great deal of effort on attacks that are unprofitable for all attackers (and thus have very low probability of happening). However, her situation is improved by the fact that exogenous fraud detection reduces her risk of harm ($1-Pr\{\text{SP}\}$). In fact, since effort from the service provider affects all attacks, while her effort must be allocated between them, it is likely that increasing $Pr\{\text{SP}\}$ has a greater influence than effort she can make against any of the attacks.

6. RELATED WORK

The question of tradeoffs in security is not a new one. Numerous authors have pointed out that, even though security is often looked at as binary, it cannot escape the budgeting, tradeoffs and compromises that are inevitable in the real world. The scalable nature of many web attacks has been noted by many authors, and indeed this has often been invoked as a possible source of weakness for attackers. Anderson [31] shows that incentives greatly influence security outcomes and

demonstrates some of the perverse outcomes when they are mis-aligned. Since 2000 the Workshop on the Economics of Information Security (WEIS) has focussed on incentives and economic tradeoffs in security.

There have been numerous studies documenting the enormous range of internet attacks. Sariou *et al.*[38] perform an interesting measurement study of internet attacks. Kanich *et al.*[27] document the result of observing a spamming botnet for a number of weeks. Their findings provide interesting insight into the scale and yield of large-scale Internet attacks. Prior to their work, we have had surprisingly little data on the cost and scale of spam campaigns. Stone *et al.*[10] also managed to take over a botnet for a period of weeks.

Varian suggests that many systems are structured so that overall security depends on the weakest-link [19]. Gordon and Loeb [28] describe a deferred investment approach to security. They suggest that, owing to the defender's uncertainty over which attacks are most cost effective, it makes sense to "wait and see" before committing to investment decisions. Boehme and Moore [32] develop this approach and examine an adaptive model of security investment, where a defender invests most in the attack with the least expected cost. Interestingly, in an iterative framework, where there are multiple rounds, they find that security under-investment can be rational until threats are realized. Unlike much of the weakest-link work, our analysis focusses on the attacker's difficulty in selecting profitable targets rather than the defender's difficulty in making investments. However, strategies that suggest that under-investment is not punished as severely as one might think spring also from our findings.

Schechter and Smith [39] examine the economics of an attacks on defensive systems deployed at large number of different locations. Their parallel attack model is similar in some respects to our threat model introduced in Section 2.1. However, their model does not include the cost of attack, instead the penalty is that an attacker risks apprehension and loss of winnings. Thus their framework is significantly different. They do not address the question of explaining missing attacks.

Grossklags *et al.*[18] examine security from a game theoretic framework. They examine weakest-link, best-shot and sum-of-effort games and examine Nash equilibria and social optima for different classes of attacks and defense. They also introduce a weakest-target game 'where the attacker will always be able to compromise the entity (or entities) with the lowest protection level, but will leave other entities unharmed." A main point of contrast between our model and the weakest-target game is that in our model those with the lowest protection level get a free-ride. So long as there are not enough of the to make the overall attack profitable, then even the weakest targets escape.

Fultz and Grossklags [29] extend this work by now making the attacker a strategic economic actor, and extending to multiple attackers. As with Grossklags *et al.*[18] and Schechter and Smith [39] attacker cost is not included in the model, and the attacker is limited mostly by a probability of being caught. Our model, by contrast, assumes that for Internet attackers the risk of apprehension is negligible, while the costs are the main limitation on attacks.

In contrast to the work of Schechter and Smith [39] and Grossklags *et al.* [18, 29] this paper is mosty concerned with better explaining observations. That is, our starting point is the gap between what the weakest-link approach predicts and the reality we see. We devote all of Section 5 to explaining observations that fall naturally from our model.

In earlier work we offered a partial explanation for why many attacks fail to materialize [21]. If the attack opportunities are divided between targeted attackers (who expend per-user effort) and scalable attackers (who don't) a huge fraction of attacks fail to be profitable since targeting is expensive. This paper extends this work and shows that even scalable attacks can fail to be economic. A key finding is that attacking a crowd of users rather than individuals involves facing a sum-of-effort rather than weakest-link defense. The greater robustness and well-known free-rider effects that accompany sum-of-effort systems form most of the explanation for the missing attacks.

Barth *et al.*[7] examine the question of reactive security, and show that it can be effective in settings where the defender does not myopically over-react to the most recent attacks. While the theoretical framework is rather different, our findings do echo this result insofar as we also explain how under-investment in security can be a sensible approach.

Odlyzko [6] addresses the question of achieving security with insecure systems, and also confront the paradox that "there simply have not been any big cybersecurity disasters, in spite of all the dire warnings." His observation that attacks thrive in cyberspace because they are "less expensive, much more widespread, and faster" is similar to our segmentation of broadcast attacks. Schneier [4] argues that "one of the important things to consider in threat modeling is whether the attacker is looking for any victim, or is specifically targeting you." In previous work we showed that phishing is subject to the tragedy of the commons reducing the return for all parties [22]. This complements the present paper in demonstrating that attackers compete with each other for finite resources.

7. CONCLUSION

John Wanamaker famously declared that "Half the money I spend on advertising is wasted; the trouble is

I don't know which half." This summarizes the Internet attacker's problem, except in the attacker's case it may be closer to 99.9999% waste. Charles(j) of course would prefer to direct all of his effort at those who have the highest likelihood of succumbing (*i.e.*, $Pr\{e_i(k)\}$ is highest) or the greatest value (*i.e.*, L_i is highest). However, target selection is costly and hard.

The threat model we propose goes some way to closing the gap between potential and actual harm. The constraint that attacks must be profitable in expectation removes a great many attacks that otherwise appear economic. It guarantees that the attacker sees a sum-of-effort rather than a weakest-link defense. It's not enough that something succeed now-and-then, or when the circumstances are right, or when all the ducks are in a row. When attacking users *en masse*, as Internet attackers do, attacks must be profitable at scale.

8. REFERENCES

[1] `http://pr.webroot.com/threat-research/cons/protect-your-computer-from-hackers-101210.html`.

[2] `http://gs.statcounter.com/press/microsoft-internet-explorer-browser-falls-below-50-perc-of-worldwide-market-for-first-time`.

[3] `http://www.trusteer.com/files/Flash_Security_Hole_Advisory.pdf`.

[4] `http://www.schneier.com/blog/archives/2010/01/32_million_jewe.html`.

[5] A. Odlyzko. Internet traffic growth: Sources and implications. *Proceedings of SPIE*, 2003.

[6] A. Odlyzko. Providing Security With Insecure Systems. *WiSec*, 2010.

[7] Adam Barth, Benjamin I. P. Rubinstein, Mukund Sundararajan, John C. Mitchell, Dawn Song, and Peter L. Bartlett. A Learning-Based Approach to Reactive Security. *Financial Crypto*, 2010.

[8] B. Ives and K.R. Walsh and H. Schneider. The Domino Effect of Password Re-use. In *CACM*, 2004.

[9] C. M. Bond and G. Danezis. A pact with the devil. 2006.

[10] Brett Stone-Gross, Marco Cova, Lorenzo Cavallaro, Bob Gilbert, Martin Szydlowski, Richard Kemmerer, Christopher Kruegel, and Giovanni Vigna. Your Botnet is My Botnet: Analysis of a Botnet Takeover. *CCS*, 2009.

[11] D. V. Klein. Foiling the Cracker: A Survey of, and Improvements to, Password Security. *Usenix Security Workshop*, 1990.

[12] R. Dhamija, J. D. Tygar, and M. Hearst. Why phishing works. *CHI*, 2006.

[13] B. Enright, G. Voelker, S. Savage, C. Kanich, and K. Levhchenko. Storm: when researchers collide. *;login*, 2008.

[14] Federal Trade Commission. Identity Theft Survey Report. 2007. `www.ftc.gov/os/2007/11/SynovateFinalReportIDTheft2006.pdf`.

[15] D. Florêncio and C. Herley. A Large-Scale Study of Web Password Habits. *WWW 2007, Banff.*

[16] Gartner. Phishing Survey. 2007. `http://www.gartner.com/it/page.jsp?id=565125`.

[17] D. Geer, R. Bace, P. Gutmann, P. Metzger, C. Pfleeger, J. Quarterman, and B. Schneier. Cyber insecurity: The cost of monopoly. *Computer and Communications Industry Association (CCIA), Sep*, 24, 2003.

[18] J. Grossklags, N. Christin, and J. Chuang. Secure or insure?: a game-theoretic analysis of information security games. *WWW*, 2008.

[19] H. R. Varian. System Reliability and Free Riding. *Economics of Information Security*, 2004.

[20] C. Herley. So Long, And No Thanks for the Externalities: The Rational Rejection of Security Advice by Users. *NSPW 2009, Oxford.*

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

[22] C. Herley and D. Florêncio. A Profitless Endeavor: Phishing as Tragedy of the Commons. *NSPW 2008, Lake Tahoe, CA.*

[23] H.R. Varian. Sytem Reliability and free Riding. *WEIS*, 2001.

[24] Imperva. Consumer Password Worst Practices.

[25] J. Franklin and V. Paxson and A. Perrig and S. Savage. An Inquiry into the Nature and Causes of the Wealth of Internet Miscreants. *Proc. CCS*, 2007.

[26] D. Kaminsky. Its The End Of The Cache As We Know It. *Black Hat Briefings*, 2008.

[27] C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson, and S. Savage. Spamalytics: An empirical analysis of spam marketing conversion. In *Proceedings of the 15th ACM Conference on Computer and Communications Security*, pages 3–14, Alexandria, Virginia, USA, October 2008.

[28] L.A. Gordon and M.P. Loeb. The Economics of Information Security Investment. *ACM Trans. on Information and System Security*, 2002.

[29] N. Fultz and J. Grossklags. Blue versus Red: Toward a Model of Distributed Security Attacks. *Financial Crypto*, 2009.

[30] N.G. Mankiw. Principles of Economics. *4-th ed.*, 2007.

[31] R. Anderson. Why Information Security is Hard. In *Proc. ACSAC*, 2001.

[32] R. Boehme and T. Moore. The Iterated

weakest-link: A Model of Adaptive Security Investment. *WEIS*, 2009.

[33] E. Rescorla. Security holes... who cares? *Usenix Security Symp.*, 2003.

[34] S. Egelman, L. F. Cranor and J. Hong. You've Been Warned: An Empirical Study of the Effectiveness of Web Browser Phishing Warnings. *CHI*, 2008.

[35] S. J. Murdoch, S. Drimer, R. Anderson, M. Bond. Chip and pin is broken. *IEEE Security&Privacy, Oakland*, 2010.

[36] S. Schechter, C. Herley and M. Mitzenmacher. Popularity is Everything: a new approach to protecting passwords from statistical-guessing attacks. *Proc. HotSec, 2010.*

[37] S. Saroiu, S. Gribble, and H. Levy. Measurement and analysis of spywave in a university environment. In *Proceedings of the 1st conference on Symposium on Networked Systems Design and Implementation-Volume 1*, page 11. USENIX Association, 2004.

[38] S. Saroiu, S. D. Gribble, and H. M. Levy. Measurement and Analysis of Spyware in a University Environment. *Proc. NSDI*, 2004.

[39] S. Schechter and M. Smith. How Much Security is Enough to Stop a Thief? In *Financial Cryptography*, pages 122–137. Springer, 2003.