A Research Agenda Acknowledging the Persistence of Passwords

Cormac Herley | Microsoft Research
Paul van Oorschot | Carleton University

Despite countless attempts and near-universal desire to replace passwords, they're more widely used than ever. The authors assert that, in many instances, passwords are the best-fit solution and suggest better means to concretely identify actual requirements and weight their relative importance in target scenarios.

> "Well, in our country," said Alice, still panting a little, "you'd generally get to somewhere else if you run very fast for a long time, as we've been doing." "A slow sort of country!" said the Queen. "Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!" —Lewis Carroll, *Through the Looking-Glass*

In the past 20 years, little progress has been made in terms of real-world impact of password research. Despite countless attempts to dislodge passwords, they're more widely used and firmly entrenched than ever. The list of new technologies, research efforts, and industry initiatives that have tried to supplant them is impressive in effort but disappointing in outcome. In this article, we consider the possible reasons in an attempt to learn from this failure.

We assert the need to better understand the loss situation (actual losses related to password compromises and the attack vectors they result from); our current data-poor state means perception drives decisions more than evidence. Password research has been far from systematic. For example, we still ask many of the same questions asked 15 or 20 years ago, and the literature is void of agreement on many issues for which consensus should be possible. We attribute this to the lack of a well-organized research agenda and systematically documented knowledge. Our goal is to promote a research agenda that both better supports passwords and allows progress.

Passwords' Resilience

Neither users nor security experts would mourn passwords' passing. For users, the main issue is usability. Major complaints are triggered by mandatory password changes (for example, every 90 days) and complex policies. Frustration increases greatly with the number of passwords that users must manage—larger portfolios of passwords increase forgetting and login errors.

Passwords' security shortcomings are many and well known. They are static in the short term and thus replayable on capture. Early attacks focused on their vulnerability to guessing and brute-force attacks. More recently, phishing and keystroke logging have

 Copublished by the IEEE Computer and Reliability Societies

allowed password harvesting on an industrial scale.[1] There are also economic problems. Agent-supported password resets are expensive. The alternative—self-service automated password resets—often rely on much weaker secondary authentication systems, such as "secret" questions,[2] which facilitated the compromise of Sarah Palin's email account.

Nonetheless, passwords have shown incredible persistence. More than seven years after Bill Gates declared "the password is dead,"[3] not only have we failed to get rid of them, but they continue to multiply as an almost universal means of Internet authentication, protecting hundreds of millions of accounts on some large sites. Two decades of stories on how urgent and imperative it is to replace them has had little impact:

- stronger alternatives and two-factor schemes are relegated to the fringes;
- sites that offer a choice of authentication mechanisms (for instance, PayPal, Blizzard, and World of Warcraft) find negligible user uptake of password alternatives; and
- authentication technologies involving biometrics and tokens,[4] client-side public-key infrastructure,[5] and graphical variations of passwords[6] have largely failed to gain mainstream deployment.

New proposals to replace passwords are offered with regularity, but expectations of success are so low that they're sometimes called "yet another authentication scheme" (YAAS). Progress on federated identity systems has been glacial, with the crowded and active offering space in 2004 noticeably quieter in 2011. There is little evidence of user adoption of OpenID,[7] and after a 1.0 release by the Eclipse Higgins Project in February 2008, there have been no major updates. Sxip Identity stopped supporting its Sxipper product in April 2011, and Microsoft announced in early 2011 that there would be no future versions of its federated client CardSpace.

There are many reasons for these failures:

- approaches that require client hardware (for example, fingerprints and smart cards) face the usual chicken-and-egg barrier;
- physical tokens are expensive, and few users aspire to carry the dozens required to replace all of their passwords;
- single sign-on schemes offer a single point of failure; and
- password managers often have poor support for roaming and inadequately studied usability.[8]

Moreover, the extra security of proposed alternatives to passwords might not always justify the cost. Organizations might prefer the devil they know in the form of current levels of fraud to an unknown devil of support costs for more complex technologies. Revocation is more complicated for stronger authentication; the self-service password resets on which many rely are no longer simple if hardware tokens are involved. Usability is another issue for stronger schemes (for instance, often longer login times). Ultimately, the enthusiasm that almost all parties show for getting rid of passwords hasn't translated into support for alternatives.

But not only have proposed alternatives failed, we've also learned little from these failures. Is federated identity a bad approach, or have the timing and incentives just been wrong? Do the many failed single sign-on initiatives teach that the whole idea is wrong, or merely that execution was flawed? Might password managers see wider adoption if roaming were better supported? Despite considerable research, execution, and deployment effort, very little has been ruled in or out. Single sign-on was an active topic of debate in the early 1990s and remains so today. There have been improvements—secure cookies, HTTPonly (which prevents cookie stealing), and tracking of IP addresses—but they're largely behind the scenes and typically augment rather than replace or simplify password-related user experience. Many things have changed beyond recognition in the past 20 years, but passwords have advanced little since the days when a 500-Mbyte disk cost US$600, thousands lined up overnight to buy copies of Windows 95, and the 1.5-megapixel Kodak DCS 420 digital camera retailed for $14,000.

Passwords, though unloved, deserve some words of praise. They have brought us this far: they are the means by which 2 billion Internet users access email, banking, social networking, and other services. They're essentially free from the service providers' viewpoint and are readily understood by users. They allow instantaneous account setup. Revocation is as simple as changing the password; those who forget their passwords can be emailed either reset links or the passwords themselves. All of this is automated and instantaneous. Passwords also allow access to users' accounts from anywhere in the world with nothing more than a simple browser. Deploying a functioning password system is relatively simple (although deployment errors are common[9]). Arguably, the Internet could not have grown to its current size and influence without them. Facebook grew from nothing to just shy of 1 million users before taking any funding. Every startup wishes to emulate that growth story, and in many cases, the only acceptable marginal authentication cost per

user is $0. While growing from one to a million users, authentication often must be free; in growing from one to 500 million users, there's seldom a good time to mandate a new, costlier user authentication system. Passwords have an impressive record of accomplishment.

Goals, Costs, and Benefits

Among security experts, there is near-unanimous agreement on the desirability of replacing passwords. Yet, this meta-goal is accepted without an understanding of what exactly is required of a replacement or what will improve once they are replaced. There's considerable confusion about the costs and benefits of replacing passwords, making it essentially impossible to effectively evaluate and compare proposals.

Poor security is obviously security experts' main concern. However, because even strong authentication technologies are vulnerable to certain attacks (for example, session hijacking involving client-end malware), more detail on exactly what is required of a replacement is essential. The US government's 2011 National Strategy for Trusted Identities in Cyberspace (NSTIC) initiative summarizes things concisely: "Passwords are inconvenient and insecure."[10] This would suggest that the implicit goal is more security and more usability (at reasonable cost). Although there's little to disagree with here, it doesn't point to a way forward. Improvements in security and usability must exceed some minimum threshold; incremental improvement in either is probably not worth the cost of disruption. A solution that answers all security concerns, provides unequivocally greater usability, and disrupts nothing seems unattainable. That many attempts have sought this suggests an over-constrained problem. In the absence of a silver bullet, we can't escape the messy work of tradeoffs.

> **Among security experts, there is near-unanimous agreement on the desirability of replacing passwords. Yet, this meta-goal is accepted without an understanding of what exactly is required of a replacement or what will improve once they are replaced.**

Confusion about Properties Needed

What properties do we actually need? Which weaknesses are unacceptable in a replacement, and which can we live with? What are the usability requirements, given that active Web users must authenticate to dozens of sites? Previous attempts to replace passwords demonstrate confusion about which threats to address.

For example, the problem of malware-infected clients has been with us for some time. Yet, many recent proposals, including OpenID and CardSpace, and most password managers offer no protection against malware-infected clients. There's confusion about whether, in a particular deployment environment, the guessing attacks of concern are online or offline. Relatively weak passwords might suffice if relevant attacks must be online, allowing other mitigation; greater strength is required if offline attacks apply.

Passwords have been with us since the earliest days of computing. The rules, policies, and best practices that govern their use have grown over time. Many organizations' policy requirements are enforced simply for compliance with security audits or industry best practices. The reasons for some requirements are poorly understood or long forgotten; in some cases, the threats underlying a policy item are no longer applicable, or it's unclear whether the policy accomplishes the design goal. Password expiration, as we discuss later, is an example in which evidence suggests the security objective isn't being achieved, despite high usability cost.[11]

The resources currently protected by passwords range from bank and brokerage accounts with significant assets to throwaway email accounts. Clearly, not all accounts in all environments have the same security needs. The objectives of different password-requesting websites vary immensely and aren't always centered on security. Passwords might be required to limit liability (if personal information is compromised), for legal reasons (some laws apply if a door is closed but not if open), to get an email address as username for contact information, or to increase the perception of a site's value. Not all users have the same needs—for celebrities, politicians, and people in the public eye, even email and Twitter accounts might require better protection than others need for banking. Not all passwords are equal; the consequences of compromise are at least as diverse as the assets they protect. Health records, employee accounts, and banking are at one end of the spectrum, where compromise can be extremely serious. Merchant and retailer accounts are closer to a middle ground; there might be an opportunity for mischief or vandalism, but the damage would likely be more limited. Personal email and social networking sites present the opportunity for inconvenience and reputation loss. Passwords that allow access to generic website

content and, for example, airport WiFi network passwords, rank lowest, protecting the site or network provider more than the user.

There is confusion as to whether we seek one solution or many. We assert that it's naive to expect that a single approach will supplant passwords in every nook and cranny into which they've forced themselves. Several or many technologies are necessary, which has advantages over a single solution. We noted earlier the problem is over-constrained in goals. The general confusion also suggests an insufficiently specified problem.

Inability to Quantify Harm

Passwords' insecurity certainly causes harm. Yet, the amount of harm password compromises cause is a subject of speculation. Most organizations reveal nothing of their losses unless compelled. Although there's no shortage of estimates, most lack a description of estimation methodology and many are produced by or for security vendors whose prime motivation isn't necessarily accuracy. In the last two years, estimates of cybercrime losses ranged over three orders of magnitude, from $560 million[12] to $1 trillion,[13] but the inconsistency inspires little confidence in any of these numbers. How bad are things, actually—how much harm does the average user suffer? Accurately predicting the benefit of replacing passwords requires accurately quantifying harm.

Harm is sometimes suffered by the user, sometimes by the site. Historically, a compromised user account might pose a serious threat to the network itself. Today, a compromised Hotmail account is inconvenient for the user and might be used to send spam, but it poses little threat of direct loss to the site or other users (although indirect damage from compromised accounts might result from their use to spread malware or "stuck in London" scams). Worst- and average-case harm can differ in severity by orders of magnitude. Gaining possession of an email password might in some circumstances allow an attacker access to a bank account. However, the average case is far less serious. Some harms are reversible, and some are not. Consumers are generally reimbursed for fraud-related monetary loss in the US,[14] but loss of privacy from leaked health records can't be repaired. Confusing the picture further, indirect harm can be many times greater than direct: money is the most obvious loss, but time, frustration, and reputation are also at stake. As with many forms of crime, online thieves might cause damage out of proportion to the money they make.

Password compromise doesn't always lead to harm. In fact, we have little idea how often one leads to the other.[15] Survey after survey finds that users ignore most security precautions, yet it seems implausible that 2 billion people would use the Internet if a majority suffered serious harm each year. The leak of 32 million RockYou user credentials hasn't been linked to any visible surge in fraud (albeit, proving such direct links convincingly can be difficult).[16] The reasons for this apparent lack of visible harm are poorly understood.

Evacuating funds from high-value accounts is nontrivial. There is evidence that many more accounts are compromised every year than can be evacuated and that money mules, not passwords, are the bottleneck resource in the cybercrime pipeline.[14] Privilege escalation (from low- to high-value accounts) might be harder than it appears. Stealing passwords and monetizing them are distinct events. It's quite possible that current systems are failing at preventing the first event but succeeding at preventing the second. When are passwords not the last line of defense, but simply one hurdle in a complex fraud-prevention apparatus? Academic researchers typically have no data on this. Back-end fraud detection at banks might catch more attempted fraud than researchers imagine. The research literature, largely assuming that passwords are the last line of defense, generally lacks discussion of back-end protection. The fraction of password compromises that leads to attempted fraud and the fraction of attempted fraud that succeeds are matters of speculation.

Finally, because riddance isn't an end in itself, what improves if we get rid of passwords? The goal, presumably, is to reduce actual and potential harm (or improve usability without reducing security). Inability to quantify harm precludes quantifying the expected improvement from alternatives. It's common for those making the case against passwords to cite impressively large fraud estimates. However, few attempt to establish how much reduction we might expect of a replacement. For example, the NSTIC document asserts that ID theft cost $37 billion in 2010[10] but is silent on how much, if any, can be attributed to passwords. This matters because displacing passwords will be costly, and no replacement will be free of vulnerabilities. It would be disappointing to incur all the cost only to find fraud levels unchanged (for example, if session hijacking was to replace keystroke logging). It would be counterproductive to mandate strong authentication for all email accounts if passwords aren't a major source of loss. Again, without quantification of harm, we proceed blindly.

Confusion about Cost

If replacing passwords was an easy proposition, it's likely that one of the many attempts would have succeeded by now. That progress has eluded us suggests the costs will be high. There will be benefits, of course,

but do they exceed the costs? Answering this is complicated by the number of stakeholders and their diversity of interests. No one actor owns the whole problem: users, Web service providers, browser vendors, software companies, government agencies, and law enforcement all have some involvement or stake. No one party can impose a solution, but several might veto solutions—for example, users resist innovations for which usability is poor.

Organizational difficulties and the alignment of incentives play a large role. OpenID provides a lesson in incentives—although many sites offer to be identifying parties, few accept the risk of disintermediation of becoming relying parties.[7] Economics might play as large a role as technology in deciding outcomes. The time and money that many organizations have sunk in passwords pose a large barrier to change. Not only is there no first-mover advantage in changing authentication systems, there's often real advantage in being last. Given the cost, confusion, training, and customer support calls that novel systems bring, it can be better to let others go first and learn from their experience. The risk of user defection might be unacceptable for Web service providers competing vigorously for traffic. Underestimating these factors can lead us to believe that proposals provide a far better cost/benefit tradeoff than is actually the case. The many failed attempts to replace passwords offer a cautionary lesson: many have asserted that promised (albeit unquantified) reduction in harm outweighs the business risks. This approach has a long history of failure, which will probably continue.

Although the research community can't quantify harm, individual companies presumably have estimates of their losses from ongoing threats. Their actions currently reveal a preference for password-related losses as opposed to the uncertainty of alternatives. To assume that they're wrong is to assume that the research community understands the business tradeoffs better than businesses do.

Finally, in segments in which the costs of replacement are greater than the benefits, improving usability might be the main driving force, with passwords persisting until a more usable alternative is found. Segments in which benefits of replacement can be shown to clearly dominate costs are good candidates for more complex solutions—but the "clear showing" isn't so easy.

Seeking Best Fit over Silver Bullets

Repeated and sustained effort has failed to uncover a silver-bullet replacement for passwords. It's time to admit that this is unlikely to change. No single alternative technology is likely to possess the combination of security, usability, and economic features that meet all goals in all situations. There's simply too much diversity in current uses of passwords and consequences when things go wrong and too many conflicting requirements, threat models, and competing stakeholder interests.[9]

Abandoning hope for a silver bullet, we should turn our efforts toward finding best-fit solutions—weighting security, usability, and economic requirements; considering the differences in account compromise severity; and weighting threats by relative likelihoods. Challenges in this requirements-driven prioritization problem include defining criteria for comparing proposed solutions and assigning weights for different elements.

Conventional security wisdom oversimplifies the story to a tradeoff between security and usability. The situation is far more complex than a one-dimensional space in which more of one implies less of the other. Indeed, if security and usability were inversely related, any attempt to increase both would be hopeless: only by reneging on the promise of better usability could security be increased. Neither is a one-dimensional quantity. For example, increasing the complexity of a password improves security against brute-force attack but does nothing against a host of others. Thus, security requirements must balance both usability and other potentially greater security requirements. Shoulder-surfing is certainly a threat but can't compromise credentials on the industrial scale that keystroke logging can. Although session hijacking is a realistic concern, authenticating every Facebook update and tweet with one-time codes seems overkill relative to the threat.

It's hard to escape the need to quantify various threats' relative likelihoods. As a thought experiment, consider a pie-chart counting all the accounts compromised in a year, divided into slices by compromise vectors (for example, keystroke logging, phishing, brute-force attacks, shoulder-surfing, and session hijacking). Although the range of attack vectors is large and growing, we have no demonstrated ability to quantify their relative likelihoods. We don't know the slice sizes—not even approximately.

Are more accounts likely to be compromised by brute-force guessing than by shoulder-surfing? Do more accounts succumb to keystroke logging than phishing? How often does cross-account password reuse lead to attack escalation? With very few exceptions, the relative success of each attack vector is unknown. Many experts have strong opinions on the importance of various attacks, but few have any data, which precludes comparing the effectiveness of would-be replacements (relative to requirements). If guessing attacks are insignificant relative to other threats, then accepting poor usability in return for highly complex

passwords is a bad bargain. If shoulder-surfing causes marginal harm, then solutions addressing it alone, neglecting other attacks, are of limited value. Because not all requirements can be met, any given proposal will meet some and not others. Thus, in the absence of the "pie-slice data" that would allow us to rank requirements, comparing alternatives to passwords is mere speculation.

This prioritization is important, unless all security requirements can be met at acceptable cost. Clearly, some threats are also less scalable than others. Although threat likelihoods will evolve, weighting attack importance per current prevalence is more useful than equal—or arbitrary—weighting.

We assert that passwords are the best fit for many (but alone, not the highest level of) authentication needs. Again, they are free (if we don't consider usability) and readily understood by users. They allow account access from anywhere in the world assuming only a simple browser. Revocation is as simple as changing passwords. Those who forget passwords can be mailed reset links or the actual passwords. Though far from ideal, this is common practice for low-value sites, for which all steps can be automated and instantaneous, including account setup. Thus, they accomplish many things that their numerous rivals can't. We might say that passwords are the worst possible authentication system, except for all the other systems.

Evaluating alternatives is hard and, to date, has been done largely in an ad hoc manner. Vendors are biased to sell products; researchers favor solutions in which they have had a hand. All parties tend to emphasize the danger of attacks for which they believe they have a cure or with which they have the most personal experience.

A Research Agenda Supporting Passwords

Our agenda includes a more systematic approach to comparing alternatives—and obtaining better pie-slice data—to better align the allocation of solution space effort to the observed harm vectors.

Ending the Belief that Passwords Are Dead

The incorrect assumption that passwords are dead has been harmful, discouraging research on how to improve the lot of close to 2 billion people who use them. Every effort should be made to correct this. Whereas vast attention, effort, and research has been spent on would-be replacements, relatively little has focused on studying plain old text passwords—how they're used and reused, how often they fail or are confused between accounts, and how to improve things. We're surprisingly ignorant on even very basic questions.

Over the years, usability has degraded, as everyone has more passwords, and policies favoring security at the expense of usability have tended to tighten. Although this might arguably be acceptable if passwords were on the verge of extinction (in which case, an increasing usability burden might even help coax users to consider alternatives), we must now acknowledge that they are not. Indeed, we believe that passwords will be with us in great numbers for the foreseeable future, including as a visible front end strengthened by complementary measures. Without better user-facing support, passwords represent a growing burden of user effort better spent elsewhere.

The incorrect assumption that passwords are dead has been harmful, discouraging research on how to improve the lot of close to 2 billion people who use them. Every effort should be made to correct this.

How poorly users are served by the current state of affairs is illustrated by the advice they receive. Logically, the relative amount of advice should be related to the threat likelihood. Although we lack the data to attach likelihoods to the individual pie-chart threats, we can reasonably conjecture that keystroke logging harvests more passwords than phishing attacks, and phishing harvests more than online brute-force attacks. Yet, the amount of advice users currently receive is in the reverse order. Users are bombarded with information on how to choose strong passwords. They receive a steady, though less extensive, stream of advice about phishing, urging them to "check the URL" (without explaining what exactly to check for) and to beware of look-alike URLs that don't match the exact spelling. As for keystroke-loggers, there's little beyond suggestions to run antivirus programs and keep software patched. Thus, it appears users receive the advice that is most easily given rather than the advice that addresses the harm they actually face.[15]

Understanding Strength and Attack Resistance

Enormous emphasis is put on coaxing users to choose strong passwords,[17] but there's no consensus on what strength various situations demand. This raises numerous questions, which we suggest the security

community has neglected to seriously consider for far too long.

First, how should we measure strength? Both info-theoretic entropy and the very crude estimates offered (with appropriate warnings) by the US National Institute of Standards and Technology (NIST) are poor measures when users choose common passwords[16]—for instance, "Pa$$w0rd" isn't particularly strong. Strength is better measured relevant to a large population of passwords, as popularity is a main determinant of risk.

Second, what strength is required to resist online attacks (assuming rate limitation)? The answer is nontrivial; it might depend on the target population's scale, as many guessing attacks are easier to conceal in the traffic of a large site. Next, how should we achieve a desired level of strength? For example, different ways of achieving the same strength can have radically different usability properties. Minimizing the usability impact of a strength requirement has seen surprisingly little work. Related, but slightly different, how should we impose a desired level of strength? Users especially dislike the policies that constrain password length and composition. Are there better means to the same end?

Third, in what scenarios are lockout or rate-limiting policies unacceptable? An argument against these policies is that they admit denial-of-service attacks. Yet, for many sites, living with this threat is preferable to imposing greater strength requirements.[9]

Fourth, when acceptable, how can lockout or rate limiting best be accomplished? By locking accounts after three failed logins? 10? or more? Is an exponentially increasing delay between attempts better than a fixed limit?

Fifth, when are offline attacks a threat? Although dependent on implementation, access to salted hashed passwords requires attacker effort. Long gone are the days when password hash files were by default world readable. A disgruntled ex-sysadmin who steals hashed passwords is the often-conjectured foe in this attack; yet, if untrusted individuals have had unfettered, unaudited access to the authentication server, a site's problems go well beyond password strength.

Sixth, are there ways to protect against offline attacks besides password strength? Mandating password changes once hashes leak might be better than strong policies at all times. Only if a leak goes unnoticed (and a password change isn't forced) does strength potentially help. Of course, reliably detecting leaks or break-ins remains difficult. Why haven't known methods extending the 20-year old ideas of Encrypted Key Exchange[18] been more widely pursued or deployed?

Finally, how much strength is required to protect against offline attacks? The bar is clearly much higher than for online attacks (assuming lockout or rate-limiting policies are in place), but at what strength are attacks effectively addressed? More strength is always better for security, but it comes at significant usability cost.

Policies and Support Tools

To address these issues, we need better policies and support tools.

Password aging policies. Password expiration policies (for example, mandating passwords be changed every 90 to 180 days) are a frequently mentioned usability disaster. They raise the cognitive burden on users, increase login errors, and lock legitimate users out of older machines and archived files. The justification of such policies applies only in a small set of scenarios: they reduce the time that an attacker has to access an account (if undetected) and the time to perform a brute-force attack on the password in the case of offline attacks. However, a study by Yinqian Zhang and colleagues found that an attacker who knew the old password could quickly guess the new one 41 percent of the time in offline and 17 percent of the time in online attacks.[11] Thus, despite their usability burden, expiration policies don't appear to deliver the intended security benefit. We suggest (as do an increasing number of security experts) that expiration policies be eliminated on the grounds that best evidence implies cost greatly exceeds benefit in all but contrived circumstances.

Realistic password guidance. Managing a large collection of passwords is a problem that most users face, but on which the research literature offers few insights or guidance. Experts now frequently challenge the historical injunction to never write passwords down as unrealistic and poor advice (obviously, it's important where the written record is stored). Users are also advised to make them strong, never reuse them, change them often, and never use them on untrusted machines. This advice is, of course, almost universally ignored. The fact that even the most conscientious users find it impossible to comply is often taken as evidence that passwords are dead and is used to support the arguments to replace them. We suggest, instead, that it's evidence of a failure by the research community to grapple with the real-world constraints of the Internet-using population. Users need realistic guidance to cope with the dozens of passwords they must now manage. Although passwords might not be viewed as the "rocket science" of security research, their scale of deployment is such that any improvement in their usability would be hard to equal for impact.

Password managers. Password managers (whether browser based, client application, or in the cloud) offer to relieve much of the cognitive burden of multiple passwords. Thus, they're potentially of great interest for scenarios in which passwords play a part in a best-fit answer. We assert that the properties of offerings in this space are largely under-studied and that development and analysis of serious password manager tools, and recognition of their potential benefits, offer great opportunities in usability and security research. Among important challenges here are security (recall that most password managers have no malware resistance) and device dependence—addressing users on machines other than their primary devices.

Prioritizing Competing Requirements

If all requirements can't be met, then some must be omitted in favor of others. The challenge is how to do so systematically rather than on an ad hoc basis. Without requirement rankings, all features have equal weight—for example, protecting against shoulder-surfing and keylogging. Again, this seems wrong because scalability implies the latter can deliver far greater harm. We've proposed that requirements be ranked in proportion to the compromises that they currently address. Although this approach is imperfect—the numbers can change as attackers adapt to defenses and evolve their techniques—using a ranking based on observed harm is preferable to choosing which threats to address arbitrarily.

There are two parts to this ranking. First, threats that currently cause significant harm must be ranked high—by definition, they have a demonstrated ability to scale. For example, if malware-infected clients result in significant credential stealing, then any solution not addressing this threat might not meaningfully reduce fraud. Second, threats that cause little observed harm require careful analysis. Some might remain dormant while more effective attacks exist; others might not scale sufficiently to harm large populations. Distinguishing these cases is important. Thus, to rank requirements, we need a much better understanding of which attacks are causing how much of the damage, or at least their relative levels. Populating the pie-chart with threat likelihoods is of first-order importance.

Agreement on a standardized, superset threat model for reference would greatly facilitate comparing solutions. This would spring naturally from the ranked list of attacks, with the highest-ranked ones forming a checklist. Rating proposals against this standard checklist would directly improve research—for example, it would immediately reveal the deficiencies of solutions that address phishing but not keylogging or brute-force attacks, or that address shoulder-surfing alone. Given the diversity of threat vectors, the limited appeal of such single-feature solutions will become obvious if we have consensus on both a superset list and a ranking of threats thereon.

We need better understanding of the harm users suffer when things go wrong. Worst- and average-case harm differ enormously. For example, by the domino effect of password reuse, a compromised low-value account might lead, worst case, to financial catastrophe. However, the almost routine leaking of millions of passwords from low-value sites (such as RockYou or Gawker), evidently with little visible effect, suggests that the average case might be very different. Partnering between the research community and data-rich organizations to facilitate data analysis is one way forward.

Finally, assuming that passwords are a best fit for many situations, it's important to segment the problem space. For those account types and situations in which passwords are likely to persist, better password support is a vast opportunity for improvement. Identifying the account types or scenarios in which passwords aren't the best fit and why (for example, when the harm is too great) is the first step to finding better alternatives.

Passwords have proven themselves a worthy opponent—all who have attempted to replace them en masse have failed. Little progress has been made in the last 20 years, in the sense that usability has degraded significantly and security hasn't improved. The reason, we suggest, is widespread confusion about why we're trying to replace them, what's required of a replacement, and what improvement is expected once they're replaced. To avoid spinning in place for another 20 years, we must do things differently. ■

Acknowledgments

We thank Kemal Biçakci, Sonia Chiasson, Serge Egelman, Markus Jakobsson, Susan Landau, Frank Stajano, and anonymous reviewers for comments that greatly improved this article. The second author acknowledges Natural Sciences and Engineering Research Council (NSERC) funding a Canada Research Chair, Discovery Grant, and NSERC ISSNet.

References

1. M. Jakobsson and S. Myers, *Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft*, Wiley-Interscience, 2006.
2. M. Just and D. Aspinall, "Personal Choice and Challenge Questions: A Security and Usability Assessment," *Proc. 5th Symp. Usable Privacy and Security*, ACM, 2009, pp. 1–11.

3. W.H. Gates III, Keynote Presentation, RSA Conference, 2004.
4. L. O'Gorman, "Comparing Passwords, Tokens, and Biometrics for User Authentication," *Proc. IEEE*, vol. 91, no. 12, 2003, pp. 2019–2040.
5. R. Housley and T. Polk, *Planning for PKI*, Wiley, 2001.
6. R. Biddle, S. Chiasson, and P.C. van Oorschot, "Graphical Passwords: Learning from the First Twelve Years," to be published in *ACM Computing Surveys*, vol. 44, no. 4, 2012.
7. S.-T. Sun et al., "A Billion Keys, but Few Locks: The Crisis of Web Single Sign-on," *Proc. Workshop New Security Paradigms* (NSPW 10), ACM, 2010, pp. 61–72.
8. S. Chiasson, P.C. van Oorschot, and R. Biddle, "A Usability Study and Critique of Two Password Managers," *Proc. 15th Usenix Security Symp.*, Usenix, 2006, pp. 1–16.
9. J. Bonneau and S. Preibusch, "The Password Thicket: Technical and Market Failures in Human Authentication on the Web," *Proc. 9th Workshop Economics of Information Security* (WEIS 10), 2010; http://weis2010.econinfosec.org/papers/session3/weis2010_bonneau.pdf.
10. "National Strategy for Trusted Identities in Cyberspace: Why We Need It," Nat'l Inst. Standards and Tech., 2011; www.nist.gov/nstic/NSTIC-Why-We-Need-It.pdf.
11. Y. Zhang, F. Monrose, and M.K. Reiter, "The Security of Modern Password Expiration: An Algorithmic Framework and Empirical Analysis," *Proc. 17th ACM Conf. Computer and Comm. Security* (CCS 10), ACM, 2010, pp. 176–186.
12. *2009 Internet Crime Report*, Internet Crime Complaint Center, 2010; www.ic3.gov/media/annualreport/2009_ic3report.pdf.
13. *Cybersecurity: Assessing Our Vulnerabilities and Developing an Effective Response*, Senate Hearing 111-43, Committee on Commerce, Science, and Transportation, 19 Mar. 2009; www.gpo.gov/fdsys/pkg/CHRG-111shrg50638/html/CHRG-111shrg50638.htm.
14. D. Florêncio and C. Herley, "Phishing and Money Mules," *IEEE Workshop Information Forensics and Security* (WIFS 10), IEEE CS, 2010, pp. 1–5.
15. C. Herley, "So Long, and No Thanks for the Externalities: The Rational Rejection of Security Advice by Users," *Proc. Workshop New Security Paradigms* (NSPW 09), ACM, 2009, pp. 133–144.
16. M. Weir et al., "Testing Metrics for Password Creation Policies by Attacking Large Sets of Revealed Passwords," *Proc. 17th ACM Conf. Computer and Comm. Security* (CCS 10), ACM, 2010, pp. 162–175.
17. J. Yan et al., "Password Memorability and Security: Empirical Results," *IEEE Security & Privacy*, vol. 2, no. 5, 2004, pp. 25–31.
18. S.M. Bellovin and M. Merritt, "Encrypted Key Exchange: Password-Based Protocols Secure against Dictionary Attacks," *Proc. IEEE Symp. Research in Security and Privacy*, IEEE CS, 1992, pp. 72–84.

Cormac Herley is a principal researcher at Microsoft Research. His current interests include the overlap among security, economics, usability, and data analysis. Herley has a PhD in electrical engineering from Columbia University. Contact him at cormac@microsoft.com.

Paul van Oorschot is a professor of computer science at Carleton University, where he is Canada Research Chair in Authentication and Computer Security. His research interests include authentication and identity management, security and usability, software security, and computer security. Van Oorschot has a PhD in computer science from the University of Waterloo. He's on the editorial board of *IEEE Transactions on Information Forensics and Security* and *IEEE Transactions on Secure and Dependable Computing* and is a member of ACM and IEEE. Contact him at paulv@scs.carleton.ca.