A Research Agenda Acknowledging the Persistence of Passwords*

Cormac Herley
Microsoft Research, Redmond, USA

Paul C. van Oorschot
Carleton University, Ottawa, Canada

Abstract

Despite countless attempts and near-universal desire to replace them, passwords are more widely used and firmly entrenched than ever. Our exploration of this leads us to argue that no silver bullet will meet all requirements, and not only will passwords be with us for some time, but in many instances they are the solution which best fits the scenario of use. Among broad authentication research directions to follow, we first suggest better means to concretely identify actual requirements (surprisingly overlooked to date) and weight their relative importance in target scenarios; this will support approaches aiming to identify best-fit mechanisms in light of requirements. Second, for scenarios where indeed passwords appear to be the best-fit solution, we suggest designing better means to support passwords themselves. We highlight the need for more systematic research, and how the premature conclusion that passwords are dead has lead to the neglect of important research questions.

1 Introduction

"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

One view of password research is that little progress has been made in the past 20 years. Despite countless attempts to dislodge them, passwords are 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, and disappointing in outcome. We consider the possible reasons in an attempt to learn from this failure. We find that despite almost universal agreement on the desirability of finding something to replace passwords, much confusion has resulted from a failure to specify both the actual requirements needed of a replacement, and a relative ranking of such requirements. If a solution which satisfies all needs cannot be found, then "best fit" approaches should be explored. The premature conclusion that passwords are dead has generated some perverse effects. We argue that it is time to admit that passwords will be with us for some time, and moreover, that in many instances they are the best-fit among currently known solutions.

We suggest two broad research directions. First, we suggest research that identifies scenarios where passwords are indeed the best fit and encourages means to better support them; this could have tremendous positive impact given the scale of password deployment. Second, we suggest research systematically prioritizing competing requirements (as rarely can all requirements be met), and using this in comparing alternatives.

We assert the need to better understand the loss situation (what the actual losses related to password compromises are, and what 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 a lack of a well-organized research agenda, and a lack of systematically documented knowledge. Our goal is to promote a research agenda that both better supports passwords, and allows progress forward.

2 The Resilience of Passwords

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

The security shortcomings of passwords are many and well-known. They are static in the short term and thus replayable upon capture. Early attacks focussed on their vulnerability to guessing and brute-forcing. More recently phishing and keystroke logging have allowed password harvesting on an industrial scale [8]. 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" [9] (*e.g.*, facilitating compromise of Sarah Palin's email account).

Nonetheless, passwords have shown incredible persistence. More than seven years after Bill Gates declared (2004) "the password is dead", 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 (*e.g.*, Paypal, Blizzard World-of-Warcraft) find negligible user uptake of password alternatives. End-user authentication technologies involving biometrics and tokens (see O'Gorman [11]), client-side public-key-infrastructure (see Housley and Polk [7]), and graphical variations of passwords (see Biddle et al. [1]), 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 are sometimes labeled as Yet Another Authentication Scheme (YAAS). Progress on federated identity systems has been glacial. A crowded and active offering space in 2004 is noticeably quieter in 2011. There is little evidence of user adoption of OpenID [12]. 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 (*e.g.*, fingerprints, smartcards) face an obvious chicken-and-egg barrier. Physical tokens are expensive and few users aspire to carry the dozens that would be required to replace all of their passwords. Single-signon schemes offer a single point of failure. Password managers often have poor support for roaming and inadequately studied usability [3]. The extra security of proposed alternatives to passwords may not always justify the cost. Organizations may 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. Self-service password resets that many rely upon are no longer simple if hardware tokens are involved. Usability is an issue for many stronger schemes (*e.g.*, longer login times). The enthusiasm that users show for getting rid of passwords has not translated into support for alternatives. Non-technical issues are possibly to blame. Mis-aligned incentives can cause desirable solutions to fail. Overall, the reasons for these failures are as many and diverse as the failures themselves.

Not only have proposed alternatives failed, but we have learnt little from the failures. Is federated identity a bad approach, or have the timing and incentives just been wrong? Do the many failed single-signon initiatives teach that the whole idea is wrong, or merely that execution has been flawed? Might password managers see wider adoption if roaming were better supported? In spite of considerable research, execution and deployment effort, very little has been ruled in or out; those who seek to replace passwords in 2012 ask many of the same questions asked in 1995. Single sign-on was an active topic of debate in the early 1990's and remains so today. No progress results from continuing to revisit the same questions without decision. There have been improvements, secure cookies, HTTPonly (which prevents cookie stealing) and tracking the history IP addresses, for example. However these have largely been behind-the-scenes and have not affected the user experience. While many things have changed beyond recognition, passwords have advanced little since the days when a 500 MByte disk cost $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 two billion Internet users access email, banking, social networking and other services. They are essentially free from the service provider 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 (this practice, though insecure, is common for low-value sites). All of this is automated and instantaneous. They allow access to one's accounts from anywhere in the

world assuming nothing more than a simple browser. Sophisticated users can protect themselves from many of the threats. Deploying a functioning password system is relatively simple (although deployment errors are common [2]). Arguably, the Internet could not have grown to its current size and influence without them. Some non-profit sites, such as Wikipedia and Craigslist, have tens of millions of users. Facebook grew from nothing to just shy of one million users *before taking any funding*. Every startup wishes to emulate that growth story, and in many cases the only acceptable marginal 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 is seldom a good time to mandate a new (more costly) user authentication system. Passwords have an impressive record of accomplishment.

3 Confusion Reigns: Goals, Costs, 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, and what will improve once they have been replaced. There is considerable confusion as to the costs and benefits of replacing passwords. This makes it essentially impossible to effectively evaluate and compare proposals.

Poor security is obviously the main concern of security experts. However, since even strong authentication technologies are vulnerable to certain attacks (*e.g.*, session hijacking involving client-end malware), more detail on exactly what is required of a replacement is essential. The U.S. government's 2011 NSTIC initiative—"National Strategy for Trusted Identities in Cyberspace" [10]—summarizes things concisely: "passwords are inconvenient and insecure." This would suggest that the implicit goal is: "more security, more usability (at reasonable cost)." While there is little to disagree with here, this does not point to a way forward. There must be minimum security and usability goals to be met; 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: it is likely impossible to find a silver bullet achieving all goals simultaneously. That many attempts have sought this suggests an over-constrained problem. In the absence of a silver bullet the messy work of trade-offs cannot be escaped.

3.1 Confusion on 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 as to which threats to address.

As one example, the problem of malware-infected clients has been with us for some time and continues. Yet, many recent proposals, including OpenID and Cardspace and most password managers offer no protection against malware-infected clients. There is confusion about whether, in a particular deployment environment, the guessing attacks of concern are online or off-line. Relatively weak passwords may suffice if relevant attacks must be online, allowing other mitigation; much greater strength is required if off-line 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. The policy requirements of many organizations 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 is unclear whether the policy accomplishes the design goal. Password expiration, as discussed in Section 5, is an example where there is a high usability cost, yet the best evidence suggests the security objective is not being achieved [15].

The resources currently protected by passwords are as diverse as the Internet itself, ranging 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 web-sites vary immensely, and are not 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 convey or increase the feeling of value in a site. Not all users have the same needs—for celebrities, politicians and people in the public eye, even email and Twitter accounts may require better protection than others need for banking. Not all passwords are equal. The consequences of compromise are at least as diverse as what they protect. Health records, employee accounts and banking are at one end of the spectrum; compromise here can be extremely serious. Merchant and retailer accounts come next; there might be an opportunity for mischief or vandalism but the damage is likely more limited. Email and social networking sites present the opportunity for inconvenience and reputation loss. Passwords that allow access to site content, airport WiFi networks *etc.*, rank lowest, protecting the site more than the user.

There is confusion as to whether we seek one solution or many. We assert that it is naive to expect that a

single approach will supplant passwords in every nook and cranny into which they have forced themselves; several or many technologies are necessary, which itself has advantages over a single solution. We noted earlier the problem is over-constrained in goals. The general confusion suggests a problem also insufficiently specified.

3.2 Inability to quantify harm

The insecurity of passwords certainly causes harm. Yet, how much harm exactly is caused by password compromises is a subject of speculation. Most organizations reveal nothing of their losses unless compelled. While there is no shortage of estimates, most lack a description of methodology, and many are produced by or for security vendors whose prime motivation is not necessarily accuracy. In the last two years, estimates of "cyber-crime losses" ranged over three orders of magnitude, from $560 million (P. Peterson, Cisco) to $1 trillion (E. Amoroso, AT&T); 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 poses little threat of direct loss to the site or other users (although indirect damage from compromised accounts may 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 to access a bank account. However the average case is far less serious. Some harms are reversible and some not. Consumers are generally made whole for direct costs upon loss of money [5]. Loss of privacy from leaked health records cannot 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 may cause damage out of proportion to the money they make.

Password compromise does not always lead to harm. In fact, we have little idea how often one leads to the other [6]. Survey after survey finds that users ignore most security precautions, yet it seems implausible that two billion people would use the Internet if a majority suffered serious harm each year. The leak of 32 million RockYou user credentials [13] has not been linked to any visible surge in fraud (albeit, proving such direct links convincingly can be difficult). The reasons for this apparent lack of visible harm are poorly understood.

Evacuating funds from high value accounts is non-trivial. There is evidence that many more accounts are compromised every year than can be evacuated and that money mules, not passwords, are the bottle-neck resource in the cyber-crime pipeline [5]. Privilege escalation (from low to high value accounts) may be harder than it appears. Stealing passwords and monetizing stolen passwords are distinct events. It is 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 defence, but simply one hurdle in a complex fraud prevention apparatus? Academic researchers typically have no data on this. Back-end fraud detection at banks may 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. What fraction of password compromises lead to attempted fraud, and what fraction of attempted fraud succeeds, are simply matters of speculation.

Finally, since riddance is not 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 is common to cite impressively large fraud estimates when making the case against passwords. However, establishing how much reduction we might expect of a replacement is rarely attempted. For example the NSTIC document [10] asserts that ID theft cost $37 billion in 2010, but is silent on how much, if any, of this can be laid at the door of passwords. This matters, as displacing passwords will be costly, and no replacement will be free of vulnerabilities itself. It would be disappointing to incur all the cost only to find fraud levels unchanged (*e.g.*, if session-hijacking were to replace keystroke logging). It would be counter-productive to mandate strong authentication for all email accounts, if passwords are not a major source of loss. Again, without quantification of the harms we proceed blindly.

3.3 Confusion on cost of ousting passwords

If replacing passwords were an easy proposition, it is likely that one of the many attempts would have succeeded by now. That progress has eluded us suggests that the costs will be large. There will also be benefits, of course. Do the benefits exceed the costs? Answering this is complicated by the number of stake-holders 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. The ben-

efit of any proposal may exceed the cost for one party, but not for others. No one party can impose a solution, but several may veto solutions; *e.g.*, users resist innovations where usability is poor.

Organizational difficulties and the alignment of incentives plays a large role. OpenID provides a lesson in incentives: while many sites offer to be identifying parties, few accept the risk of disintermediation of becoming relying parties [12]. Economics may play as large a role as technology in deciding outcomes. The sunk costs that many organizations have in passwords pose a large barrier to change. Not only is there no first-mover advantage, in moving to any new authentication system, there is often real advantage in being last. Given the cost, confusion, training and customer support calls that introducing a novel system brings it can be better to let others go first and learn from their experience. The risk of user defection may be unacceptable for web service providers competing vigorously for traffic. Underestimating these factors can lead us to believe that proposals have far better cost/benefit tradeoff than is actually the case. The many failed attempts to replace passwords offer a cautionary lesson: asserting that promised (un-quantified) reduction in harm outweighs the business risks has been tried many times. It has a long history of failure, and this will probably continue.

While the research community is unable to 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 trade-offs better than businesses themselves do.

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

4 Seeking Best-Fit over Silver Bullets

Repeated and sustained effort has failed to uncover a silver bullet replacement for passwords. It is 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 meets all goals in all situations. There is 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 [2].

Abandoning hope for a silver bullet, we should turn our efforts towards finding best-fit solutions—by the messy work of 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.

While conventional security wisdom oversimplifies the story to a trade-off between security and usability, the situation is far more complex than a one-dimensional space where more of one implies less of the other. Indeed if they 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 attacks, but does nothing against a host of others. Thus, security requirements must be balanced against both usability and other potentially greater security requirements. Shoulder-surfing is certainly a threat, but is entirely incapable of compromising credentials on the industrial scale that keystroke logging can. While session hijacking is a realistic concern, authenticating every Facebook update and tweet with one-time codes seems overkill relative to the threat.

As a tool for ranking properties it is hard to escape the need to quantify the relative likelihoods of various threats. As a thought experiment consider a pie-chart counting all the accounts compromised in a year, divided into slices by compromise vectors (*e.g.*, keystroke logging, phishing, brute-forcing, shoulder-surfing, session hijacking, ...). While the range of attacks 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 re-use lead to attack escalation? Sadly with very few exceptions, the relative success of each attack vector is unknown. Many have strong opinions on the importance of various attacks, but few have any data. This 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, while neglecting other attacks, are of limited value. Since 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

reduces to speculation.

Identifying the threat vectors is easy work, compared to the important task: determining their relative likelihoods and impact. This prioritization is important, unless all security requirements can be met (at acceptable cost). Clearly some threats are also less scalable than others. While threat likelihoods will evolve, weighting attack importance per current prevalence is more useful than equal (or arbitrary) weighting of all attacks.

We assert that passwords themselves are the best-fit for many (but not the highest level) authentication needs. 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 set-up. Thus passwords accomplish many things that their numerous rivals cannot. Indeed it might be said of passwords that they 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 which they have most personal experience with. Our agenda suggested in Section 5 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.

5 A research agenda supporting passwords

Building on the above discussion, we seek to promote a research agenda better supporting passwords. We also highlight research questions—some long overdue and neglected—that we believe deserve attention.

5.1 Ending belief 'Passwords are dead'

The spectacularly incorrect assumption "passwords are dead" has been harmful, discouraging research on how to improve the lot of close to two billion people who use them. Every effort should be made to correct this. While vast attention, effort and research has been spent on would-be replacements, there has been relatively little on studying plain old text passwords themselves: how they are (re)used, how often they fail or are confused between accounts, and how to improve things. We are surprisingly ignorant on even very basic questions.

During this time usability has degraded: everyone has more passwords, and policies have tended to tighten rather than loosen over time. While 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 that is better spent elsewhere.

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. While we cannot attach likelihoods to the individual pie-chart threats of Section 4 it is reasonable to conjecture that keystroke logging harvests more passwords than phishing attacks, and phishing harvests more than online brute-forcing. 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 look-alike URLs which don't match the exact spelling. As for keystroke-loggers there is little beyond suggestions to run anti-virus programs and keep software patched. Thus, they receive, it appears, the advice that is most easily given, rather than the advice that addresses the harms they actually face [6].

The above reiterates the need for data on "pie slice sizes" on which to base advice to users, and more generally, to expend greater research effort on understanding problems related to text password themselves.

5.2 Understanding strength, online, offline

Enormous emphasis is put on coaxing users to choose strong passwords [14]. Yet there is 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 strength be measured? Info-theoretic entropy and the NIST criteria are poor measures [13] when users choose common passwords, *e.g.*, '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 in place)? The answer is non-trivial; it may depend on the scale of the target population, as many guessing attacks are easier to conceal in the traffic of a large site. Next, how should a desired level

of strength be achieved? For example, different ways of achieving the same strength can have radically different usability properties. The question of minimizing the usability impact of a strength requirement has seen surprisingly little work. Related, but slightly different, how should a desired level of strength be imposed? The policies that constrain password length and composition appear especially hated by users. 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 [2].

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

Fifth, when are off-line attacks a threat? While 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 un-trusted 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 off-line 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 itself remains difficult.

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

5.3 Better policies and support tools

Password aging polices. Password expiration policies (*e.g.*, mandating passwords be changed every 90–180 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 brute-force the password in the case of off-line attacks. However, a study by Zhang *et al.* [15] found that an attacker who knew the old password could quickly guess the new one 41% of the time with an off-line or 17% of the time with an online attack. 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. The historical injunction to never write passwords down is now frequently challenged by experts as unrealistic and poor advice (obviously, it is important where the written record is stored). Users are also advised to make them strong, never re-use, change them often, and never use them on un-trusted 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 is evidence of a failure by the research community to grapple with the real-world constraints of the Internet-using population. Rather than advice that is bound to be ignored, users need realistic guidance to cope with the dozens of passwords they must now manage. While passwords may not seem "hot" research, the 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 are potentially of great interest for scenarios where passwords are the best-fit answer. We assert that the properties of offerings in this space are largely understudied, 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 itself (recall that most password managers have no malware resistance), and addressing roaming users (i.e., users using machines other than their primary devices).

5.4 Prioritizing competing requirements

If all requirements cannot 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. If we have no ranking of requirements then all features have equal weight—such as, e.g., protecting against shoulder-surfing and keystroke-logging. This seems wrong, as scalability implies the latter can deliver far greater harm. We have proposed that requirements be ranked in pro-

portion to the compromises that they currently address. While 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 in an arbitrary way. 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 may not meaningfully reduce fraud. Second, threats that cause little observed harm require careful analysis. Some may remain dormant while more effective attacks exist; others may 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 of Section 4 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, this would immediately reveal the deficiencies of solutions that address phishing but not keylogging or brute-forcing, 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 a ranking of threats.

We need better understanding of the harms suffered by users when things go wrong. Worst-case and average case harm differ enormously. For example, by the domino effect of password re-use, a compromised low-value account *might* lead to financial catastrophe for a user. However, the almost routine leaking of millions of passwords from low-value sites (*e.g.*, RockYou and Gawker), evidently with little visible effect, suggests that the average case may 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, then it is important to segment the problem space. For those account types and situations where passwords are likely to persist, supporting passwords better is a vast opportunity for improvement. Identifying the account types or scenarios where passwords are not the best-fit and why (*e.g.*, when the harm is too great) is the first step to finding better alternatives.

6 Concluding Remarks

Passwords have proved themselves a worthy opponent: all who have attempted to replace them have failed. It is fair to say that little progress has been made in the last 20 years: usability has degraded significantly, while security has not improved. The reasons, we suggest, are widespread confusion about why we are trying to replace them, what is required of a replacement, and what improvement is expected once they are replaced. To avoid spinning in place for another 20 years, we must do things differently.

The "password replacement problem" is both underspecified (in vagueness of specific goals or concrete requirements) and over-constrained (it is simply impossible to find a single solution addressing all security and usability needs in all scenarios). Regarding the latter, rather than seeking a silver bullet, we must consider best-fit solutions. We offer two major conclusions.

First, we assert that passwords are themselves the best-fit for many of the scenarios in which they are currently used. No other single technology matches their combination of cost, immediacy and convenience that many scenarios require; they are likely to persist for some time. The research avenue this motivates is exploring how to better support the use of passwords. Second, there are scenarios where passwords are not the best-fit. To determine these cases and find suitable alternatives requires (a) clearly and specifically identifying the requirements; and (b) devising a practical, reliable methodology to evaluate, score, and compare competing alternatives against these requirements. The research avenue suggested is to gain better insight into actual threat likelihoods and actual harms experienced when things fail. An important step is to replace unsupported opinions by factual evidence and real-world data.

Acknowledgements. We thank Kemal Biçakci, Sonia Chiasson, Serge Egelman, Markus Jakobsson, Susan Landau, Frank Stajano and anonymous reviewers for comments that greatly improved this paper. The second author acknowledges NSERC funding a Canada Research Chair, Discovery Grant, Discovery Accelerator Supplement, and NSERC ISSNet.

References

[1] R. Biddle, S. Chiasson, P.C. van Oorschot. Graphical Passwords: Learning from the First Twelve Years. ACM Computing Surveys vol.44 no.4 (to appear).

[2] J. Bonneau and S. Preibusch. The password thicket: Technical and market failures in human authentication on the web. WEIS 2010, Cambridge, MA, USA.

[3] S. Chiasson, P.C. van Oorschot, and R. Biddle. A usability study and critique of two password managers. In *15th USENIX Security Symposium*, pages 1–16, 2006.

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

[5] D. Florêncio, C. Herley. Phishing and Money Mules. IEEE Workshop on Information Forensics and Security (WIFS 2010).

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

[7] R. Housley, T. Polk. *Planning for PKI*. Wiley, 2001.

[8] M. Jakobsson, S. Myers. *Phishing and countermeasures: understanding the increasing problem of electronic identity theft.* Wiley-Interscience, 2006.

[9] M. Just and D. Aspinall. Personal choice and challenge questions: a security and usability assessment. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, pages 1–11. ACM, 2009.

[10] NIST. National Strategy for Trusted Identities in Cyberspace. Why We Need It. `http://www.nist.gov/nstic/NSTIC-Why-We-Need-It.pdf`. 2011.

[11] L. O'Gorman. Comparing passwords, tokens, and biometrics for user authentication. Proc. IEEE 91(12):2019-2040, Dec. 2003.

[12] S.-T. Sun, Y. Boshmaf, K. Hawkey, K. Beznosov. A billion keys, but few locks: the crisis of web single sign-on. NSPW 2010.

[13] M. Weir, S. Aggarwal, M. Collins, H. Stern. Testing Metrics for Password Creation Policies by Attacking Large Sets of Revealed Passwords. ACM CCS 2010.

[14] J. Yan, A. Blackwell, R. Anderson, A. Grant. Password memorability and security: empirical results. IEEE Security and Privacy 2(5):25-31, 2004.

[15] Y. Zhang, F. Monrose, M.K. Reiter. The Security of Modern Password Expiration: An Algorithmic Framework and Empirical Analysis. ACM CCS 2010.