The password thicket: technical and market failures in human authentication on the web

Joseph Bonneau
Computer Laboratory
University of Cambridge
jcb82@cl.cam.ac.uk

Sören Preibusch
Computer Laboratory
University of Cambridge
sdp36@cl.cam.ac.uk

Abstract

We report the results of the first large-scale empirical analysis of password implementations deployed on the Internet. Our study included 150 websites which offer free user accounts for a variety of purposes, including the most popular destinations on the web and a random sample of e-commerce, news, and communication websites. Although all sites evaluated relied on user-chosen textual passwords for authentication, we found many subtle but important technical variations in implementation with important security implications. Many poor practices were commonplace, such as a lack of encryption to protect transmitted passwords, storage of cleartext passwords in server databases, and little protection of passwords from brute force attacks. While a spectrum of implementation quality exists with a general correlation between implementation choices within more-secure and less-secure websites, we find a surprising number of inconsistent choices within individual sites, suggesting that the lack of a standards is harming security. We observe numerous ways in which the technical failures of lower-security sites can compromise higher-security sites due to the well-established tendency of users to re-use passwords. Our data confirms that the worst security practices are indeed found at sites with few security incentives, such as newspaper websites, while sites storing more sensitive information such as payment details or user communication implement more password security. From an economic viewpoint, password insecurity is a negative externality that the market has been unable to correct, undermining the viability of password-based authentication. We also speculate that some sites deploying passwords do so primarily for psychological reasons, both as a justification for collecting marketing data and as a way to build trusted relationships with customers. This theory suggests that efforts to replace passwords with more-secure protocols or federated identity systems may fail because they don't recreate the entrenched ritual of password authentication.

WEIS 2010
The Ninth Workshop on the Economics of Information Security

Contents

1 Introduction

Password authentication remain ubiquitous on the web, despite over thirty years of research demonstrating its weaknesses [80]. Countless improvements have been proposed to improve password security or replace it altogether, but none has seen any significant adoption in the market for human authentication by Internet sites. The security economics community has begun to ask hard questions about why it is so difficult to deploy better techniques [56, 83]. It seems clear that security researchers have failed to fully understand the incentives in the market for password-based authentication.

The demand side of the market is relatively well-known, with a large number of research studies documenting how users choose passwords and how they cope with the difficult requirement of maintaining passwords with many online accounts. Many users still choose easily-guessable passwords write them down, share them casually with friends, and rarely change yet frequently forget them. Most critically, users frequently re-use passwords, with the average password being shared by at least 5 sites [37].

Yet such practices cannot be written off as evidence of user ignorance or apathy. Consumer research shows that security remains the primary stated concern of e-commerce customers [74]. Most users generally understand that there are risks of using easy-to-guess passwords or re-using passwords and recognise that they should separate high-security accounts from low-security ones [45]. However, users simply have too many accounts to manage securely, with the average user holding over 25 separate password accounts [37]. Users frequently state that they re-use passwords knowing it is risky because they simply feel unable to remember any more [84], and evidence suggests users are stretching their memory to its limits: traffic logs indicate that more than 1% of all Yahoo! users forget their passwords in any given month [37], and a laboratory study showed that users are unable to remember their own passwords for as many as a quarter of sites they have registered with [45].

Due to this high level of password re-use any site's insecure practices may directly harm another site if they have overlapping user bases in a domino effect [60]. This risk has become quite salient in the past year, when a hacker was able to expose a database of 32 million email-address/password pairs from the gaming website RockYou [110], claiming that at least 10% of the credentials could be directly used to access PayPal accounts. Such cross-site account compromise attacks are said to be a common threat [59], and indeed in January 2010 Twitter forced millions of users to change their passwords after observing attacks using credentials stolen from a torrent website [89].

Password security on the web has become a complex and interconnected system, but it remains poorly understood how sites choose a security policy. We have conducted the first large-scale study of password implementations, collecting data from 150 websites about technical details of password collection, use, and reset mechanisms Our findings, documented in Section 4, confirm the widespread occurrence of questionable design choices, inconsistencies, and indisputable mistakes, suggesting that implementers perform just as poorly as users at maintaining good password practices.

Our analysis in Section 5 suggests reasons why many sites, particularly in the online news market, offer password accounts when there is little security reason for doing so. Password collection appears to be driven in this segment by a desire to collect email addresses and marketing data. Password registration may also serve as a ritual to establish trust with users and offer a feeling of membership in an exclusive group. We believe the psychology of password registration is a key avenue for future research; it has already been shown that password sharing between individuals, particularly couples, serves to reinforce a sense of intimacy [103, 32].

Our data leads us to economic explanations for why password security is failing, as introduced in Section 6. Sites' decisions to collect passwords can be viewed as a tragedy of the commons, with competing sites collectively depleting users' capacity to remember secure passwords. Lower security sites have real disincentives to implementing more security but bear no cost if their insecurity leads to compromise at higher-security sites, making password insecurity a negative externality.

Our conclusions (Section 7) are disheartening for the prospects of technological solutions to the web's password problems. Market forces oppose the introduction of single sign-on systems such as OpenID [92], and years of habit may hinder the possibility of deploying stronger authentication methods.

2 Password history and related work

2.1 Password distributions and cracking

Robert Morris and Ken Thompson published the first academic paper on password security in 1979 [80]. They presented empirical analysis of users' password choices by conducting dictionary attacks on a real system, recovering 81% of the 3,000 users' passwords using a modest dictionary. Many users picked extremely short strings as password security was not widely understood. Feldmeier and Karn followed up this analysis 10 years later with further experiments and found that they were still able to break 30% of system passwords with small dictionaries, which was considered a major improvement [36]. Later experiments by Klein (25% of passwords cracked in 1990) [66], and Spafford (20% of passwords cracked in 1992) [104], appeared to show continued improvement in user password choice. The web appears to have altered this trend. A 2006 password leak from MySpace revealed slightly weaker passwords [101], and a 2007 study by Cazier and Medlin cracked 99% of passwords at a real e-commerce website given unlimited guessing access [27].

Simultaneously, password guessing techniques have gotten much more sophisticated. John the Ripper has been developed as a general-purpose password-cracking library for over a decade and now contains dictionaries of millions of known passwords [1]. Oechslin introduced rainbow tables in 2003 as a much-improved time-memory trade-off for password-cracking [85]. These approach was later improved by Naraynana et al. to take into account real password statistics [82]. Recent research by Weir et al. demonstrates that base password dictionaries can be made stronger by automatically generating modifications based on observed models of human password choice [114].

2.2 User password management

In a different tack from the arms race of password cracking, Adams and Sasse inspired a large amount of research into corporate employees' password behaviour with their 1999 study that observed users writing down passwords and easily yielding them to social engineering attacks [13]. Further research by Sasse et al. confirmed that many users within corporate environments were overwhelmed by security advice and requirements which placed a significant strain on IT help desks [96]. Brostoff proposed as a solution to this problem that users should be given more than 3 chances to remember their passwords to avoid costly manual resets [23].

Dourish and Grinter in 2004 focused towards users' inability to manage password requirements for their personal computing needs [34]. In 2006, Riley conducted a large user study which found that most users heavily re-used a small number of web passwords between accounts and never changed them [93]. Gaw and Felten published a large user study the same year which found similarly high levels of password re-use [45]. They also brought users into the laboratory and observed them attempting to log in to all websites they had accounts with, and found that users were unable to log in to a significant number accounts due to forgotten passwords, and that password re-use was a growing problem as users registered more accounts. A landmark study in 2007 by Flôrencio and Herley provided large-scale data collected in-situ by a browser toolbar [37]. Their numbers were striking: the average web user was found to maintain 25 separate password accounts, with just 6.5 passwords. They also suggested that up to 1% of a large website's users will perform a password reset in a given month.

In addition to password re-use and forgetfulness, a number of other practices have shown up consistently in user studies. A sizable portion of users write passwords down, never change them, and based them on personally identifiable information. Awareness and adoption of automated techniques to store and enter passwords remain extremely low [93]. Password sharing between individuals is a surprisingly common practice. Singh et al. conducted a survey focused on this phenomenon in 2007, finding the practice to be very common for a variety of reasons, in particular intimacy between couples, the majority of whom acknowledged sharing passwords [103]. Password sharing for social networking sites has also been identified by boyd in her ethnographic work as a major trend amongst teen users [32].

Finally, recent work has looked into users' mental model of password security. While users generally

have skewed perceptions of the guessability of their passwords [45], there is evidence that most users recognise the difference in security requirements between sites, and attempt to segregate passwords low-security sites [84]. A recent user survey by Shay et al. at examined user attitudes towards the imposition of stringent new password requirements for accounts on Carnegie Mellon University's academic network [102]. While users were initially annoyed by the new requirements, most felt remembering more complicated passwords was worthwhile to improve security.

2.3 Improved cryptographic protocols

Morris and Thompson's seminal paper [80] described the initial implementation of password security on UNIX, including the storage of salted and hashed passwords in the `/etc/passwd` file which has become standard practice in operating systems to reducing the risk of password file compromise. Feldmeier and Karn suggested improvements in 1989 [36], including the use of secret salt values and a shadow password file to make attacks more difficult, but the basic UNIX storage protocol has remained.

Lamport suggested one-time passwords in 1981 to eliminate the risk of password theft when passwords were transmitted over insecure channels (or into insecure computing devices) [72]. Lamport's scheme used repeated hashing of an initial secret to generate sequential one-time passwords, eliminating the need for the server to store all one-time passwords. Lamport's protocol was developed by Haller into S/KEY [52] which was eventually standardised with an RFC as the OTP protocol [53]. Variants were later developed by Rubin [95] and Kuhn [70] which eliminated the risk of theft of the of the initial secret in Lamport's scheme by generating independent one-time passwords. One-time password protocols have seen limited use by general-purpose websites.[1]

An alternative to one-time password schemes is password-authenticated key exchange, where two parties can remotely establish a shared key based on knowledge of a password without transmitting the password itself. Bellovin and Merrit proposed the Encrypted Key Exchange protocol in 1992 [18]. This protocol was balanced in that both parties were assumed to know the secret password. Augmented EKE, a more applicable protocol for client-server scenarios on the Internet was introduced in 1993 [19], preventing password loss if the server is compromised. A large number of EKE variants have been proposed since, including protocols provably secure without random oracles [65, 46]. The most readily applicable variant for Internet authentication is the Secure Remote Password protocol [117], which is efficient and prevents dictionary attacks by an eavesdropper. Despite being standardised by an RFC for use with TLS [107], SRP has seen little usage at common websites.

2.4 Improving password entry

Several approaches have been taken to improve memorability and security of textual passwords. Haga and Zviran found in 1991 that cued-recall systems, where a user is asked a question and then types in a password, enabled users to remember harder-to-guess passwords [50]. This approach was expanded by Pond et al. in 2000 into a higher-security multi-word system, but over a third of study participants forgot their passwords after two weeks, but 8% were guessable by attackers [88]. A cheaper alternative is mnemonic passwords based on compressing a memorable sentence into a short password, which were found in a real-world study in 2000 to significantly reduce guessability without affecting recall [118]. However, later research found that mnemonic dictionaries based on databases of quotations, film and song lyrics were effective at guessing mnemonic passwords [71]. A simpler approach is to persuade users to pick more secure text passwords. This was argued to be the most cost-effective technique in 2001 [115], and Conlan et al. demonstrated that graphical indicators made a significant difference in password strength [29].

Because human memory is evolved to remember experiences and emotions and not random strings, graphical passwords have been proposed as a means to allow humans to efficiently enter more-difficult-to-guess secret information. Valentine proposed Passfaces in 1998, enabling users to remember a PIN by

[1] A variant of one-time passwords known as TAN codes has been deployed by some European banks, but not by websites offering free accounts.

remembering faces instead of digits [109]. Jermyn et al. introduced several new schemes in 1999 [61], some involving humans drawing a password, and other selecting a password's characters from a visual grid. Wiedenbeck et al. proposed PassPoints in 2005, which authenticates a user by having them click on secret spots in an image [116]. There have been a number of other proposals over the years, but enthusiasm in the security community has dampened as effective attacks have been shown against most graphical systems. Davis et al. found that users' choice in the Passfaces system was predictable enough to render it completely insecure [33], while Thorpe et al. found that user tendencies in other graphical password schemes reduces the effective security to far below the theoretical level and may be similar to that provided by text passwords [108].

Even secure graphical passwords are vulnerable to phishing attacks, so cognitive passwords have been proposed instead which require a human to actively compute a response to a random challenge, hiding the user's complete secret from an eavesdropper [58]. Weinshall proposed a practical scheme in 2006 [113], but it was broken within a year [47]. To date, a practical and secure scheme has not been found, and Coskun and Herley argued from an information theoretic-perspective that the limits of human memory will prevent cognitive schemes from every being both usable and secure [30].

2.5 Single sign-on and related systems

Single sign-on systems allow a user to register one password with a trusted server which can be used to access any online entity which is willing to rely on the user's chosen authentication service. The best-known historic predecessor is the Kerberos protocol [106], developed for MIT's Project Athena in the 1980's and eventually standardised in an RFC [67]. Kerberos uses symmetric encryption to authenticate a user to a trusted key server, which then provides a shared session key.

While Kerberos has been standardised as a cipher suite for use with TLS [78], it is more commonly deployed on the web using cookies and HTTP redirection. Many authentication protocols loosely based on the Kerberos model have arisen at various universities, such as the Central Authentication Service at Yale University [76], WebAuth at Stanford University [100], Raven/ucam-webauth at the University of Cambridge [112], CoSign at the University of Michigan [31], and Pubcookie at the University of Washington [3]. An overview of the different design dimensions of these and dozens of other single-sign on systems is provided in [87], most however, are designed with a single trusted authentication server in mind, a strong notion of user identity and a secure channel to distribute initial credentials, none of which are possible on the wider Internet.

The Shibboleth protocol [79] is designed to facilitate interoperation between such systems by allowing multiple identity providers, but has still mostly seen deployment at academic websites. A similar protocol which has seen some commercial adoption is OpenID [92], designed to allow any web server to rely on any other to authenticate a user. OpenID's specification is maintained by a non-profit foundation, the current version 2.0 of the protocol is a draft standard [5]. While it has been criticised for leaving users susceptible to phishing [73], it has received official support from a number of large websites. The OAuth protocol [54] is a related but distinct extension to allow websites to transfer user data directly from server to server with user authentication, it is also currently a draft standard [91].

Meanwhile, there have been several proprietary attempts at single sign-on. The most prominent was Microsoft Passport [10], introduced in 1999 and designed as a single-sign on for the entire web. Passport received criticism for its centralised nature, as Microsoft servers handled all authentication requests, and several security problems were found with the protocol [68, 86]. Within the past year, Facebook Connect [8] has grown significantly in popularity. In addition to authentication, Facebook Connect provides a set of APIs for third-party sites to request user data from Facebook and publish user activities to Facebook user's profiles. Finally, Verified by Visa [12] is a widely-deployed single sign-on system in the online payments space. It has slightly different goals as it only aims to provide supplemental assurance that an online shopper is the valid holder of a payment card, it is not designed for authentication for other purposes. It has also seen significant criticism for security shortcomings [81].

Beside cross-organisational single sign-on systems, corporate mergers and acquisitions have resulted in consolidation of user accounts and password implementations under the umbrella of a few large-scale

Web companies. Google accounts, Microsoft Live accounts, and Yahoo! accounts were introduced at absorbed Web applications as diverse as online word processing (now Google Documents, previously Writely) or photo sharing (flickr), where they replaced previous standalone login systems.

2.6 Automated Password Management

Recognising the difficulty of deploying single sign-on schemes, automated password management systems seek to achieve the same effect by have a trusted delegate perform log-in on behalf of a user, without any changes to the relying server. Gabber et al. proposed an initial scheme called Janus in 1997 which would accomplish this by proxying web traffic through a paid, trusted anonymising proxy which would automatically fill strong passwords into web forms for the user when needed [43]. Such a service never materialised, but password management by the operating system has been widely implemented, at least since Apple's KeyChain software debuted for MacOS 8.6 in 1999. There are now many free software programs to automatically store passwords [11, 9].

Most modern browsers, along with some desktop software such as email clients, will automatically remember previously entered passwords and optionally secure them through a master password. This solution adds little for security, though, as users must still create the passwords initially. Advanced users may install add-ons to generate secure passwords but these require skilled manual intervention and trust in the tools. A 2006 user survey indicated that 93% of users have never used any automated tools, although about two-thirds did make use of a browser with automatic password entry [93].

Academic research by Halderman et al. focused on building browser extensions to automatically derive domain-specific passwords by hashing a master password with the current domain [51]. A similar approach was taken by Ross et al., who also supported the difficult case of remote access to domain-specific passwords when needed through a secure server [94], their PwdHash browser extension has been installed by about 100,000 Mozilla Firefox users. Florêncio and Herley have re-visited the trusted server approach, proposing to use trusted proxy servers to securely access websites even from highly untrusted computers by using one-time password schemes to authenticate the proxy and having the proxy then authenticate with the desired web server [55, 38].

2.7 Password implementation standards

To the best of the author's knowledge, there does not exist a standard which provides direct recommendations for password implementation at commercial websites. There do exist a variety of government-published standards, but these are intended to provide guidance for government agencies and are not aimed at commercial sites. In the United States, the National Institute of Standards and Technology published FIPS 112 "Password Usage Guidelines" in 1985, which gave very specific advice on password implementations [2]. The standard includes encrypting all passwords during transmission and storage, forcing users to change passwords every year, and forcing users to pick a new password after forgetting the previous one. This advice pre-dates the consumer Internet, and some of the standard has been specifically recommended against by more recent academic work [13, 39]. On the issue of password requirements, FIPS 112 is provides three possible levels. A six character minimum is required for high-security passwords, but no non-alphabetic characters are required, though users must be directly warned not to pick dictionary words or information relating to them personally.

More recently, NIST published FIPS 800-63 "Electronic Authentication Guideline" in 2005 [26]. This document contains fewer specifics about password authentication, instead defining four assurance levels for remote authentication, for which only the lowest two levels can be based solely on passwords. It provides general requirements for password security, such as encryption during transmission and storage, but is not intended for website developers and provides no guidance on the interaction of password authentication with web protocols. The standard also avoids recommending any specific password requirements, but does mandate that the lowest level of security prevents an attacker from having more than a 2^{-10} chance of guessing a password, and the second level no greater than a 2^{-14} chance. It leaves

the achievement of this up to implementers, but provides a detailed algorithm for approximating the entropy password under different sets of requirements. For example, it estimates a 6-character user-chosen password with no restrictions will have 14 bits of entropy, which it deems sufficient to meet the lowest level of security.

Amongst international standards, ISO 27001 was published in 2005 and mandates that a secure system require strong passwords, but provides no technical details [4]. The German BSI (Federal Office for Information Security) published technical standards for satisfying ISO 27001 which do contain specific recommendations [25, 24]. In particular, the requirements include a minimum length of seven characters, a requirement of numbers or symbols, and proactive checking against a dictionary of weak passwords. Interestingly, the standard acknowledges that password change requirements are onerous and recommends against them.

None of these standards deals with the issues of password reset by email, assuming instead that an administrative help desk is available for reset. Furthermore, while recommendations against guessing attacks are made, there is no advice for how to implement this in a web server, nor are there recommendations for other details such as preventing probing for user accounts.

2.8 Empirical studies of web security

A small number of previous studies have used a similar methodological approach of empirically studying security practices at websites "in the wild." In most cases, the number of security errors discovered is considered surprisingly large to the academic security community. Fu et al. studied web authentication in 2001, specifically the format of cookies used by common websites, and found a number of flaws in session cookie generation [41]. Their paper also contained a wealth of practical advice on password implementation specifically for the web which is lacking in government standards, including a protocol for session cookie generation. We ignore in our survey the practice of generating secure cookies, as reverse-engineering the format of cookies used by a given website is a time-consuming manual task.

In 2009, three large-scale studies of website privacy practices were published, both of which involved examining a large number of randomly chosen websites. Gomez et al. studied privacy practices at 100 top web-sites, confirming the wide-spread use of privacy-tracking bugs [48]. Bonneau and Preibusch studied 45 social networking sites [21], examining the privacy practices as well as several security-relevant practices, such as TLS deployment (which was found to be very low). Krishnamurthy and Wills studied social networks and identified common practices which leak user data to third-party ad networks [69], such as including user IDs in URLs which are then transmitted as part of the HTTP `referer` header.

Two previous studies have examined security practices at banking sites. Mannan et al. signed up for five Canadian banks, but were limited to a qualitative evaluation due to the low sample size [75]. Falk et al. studied several hundred banking sites, similar to the depth we will aim for, but did not create accounts and thus only collected a few pieces of data [35]. Important findings from Falk et al.'s study were that most banking websites (76%) suffered at least one noticeable design flaw of the 5 checked for, including 30% of banks failing to use TLS, and 31% sending sensitive information via email.

Most relevant to our password study, Furnell examined the password policies of 10 major web sites in 2007 [42]. Furnell found unique advice and policies at every site studied, with the only common thread being a six-character minimum password being required at 8 sites Most site in Furnell's study had no other password requirements and failed to implement dictionary checks.

3 Methodology

3.1 Research questions

Our main goal is to further the understanding of how and why real-world websites implement password authentication. We are driven in particular towards several research questions:

RQ1 How does the user experience vary from site to site? Past user studies have suggested that user confusion and frustration can be a cause of insecure behaviour [13] and that websites have disparate password policies [42]. We seek to analyse how the user experience varies across a much larger sample of sites. We also aim to determine how current interfaces compare to the published government standards for password authentication (§ 2.7).

RQ2 What implementation weaknesses exist? Past studies of security practices on randomly selected websites have revealed many flaws, such as generating login cookies insecurely (§ 2.8). We aim to assess which weak practices are common in password systems, and how frequently they occur. Furthermore, we will attempt to assess which insecure practices are deliberate choices in favour of usability, and which can be considered to be mistakes on the part of implementers. For example, not implementing TLS may be a deliberate choice to save cost or complexity on both the client or the server side. But implementing TLS for some password submission mechanisms but not others can be assumed to be an oversight.

RQ3 Which circumstantial factors affect sites' password implementation choices? By studying a wide variety of sites with different characteristics, we hope to find out which factors may affect password practices. We can correlate observed practices with information about the size, age, traffic volume, market segment and features offered as indicators of what forces are driving security choices. We can also correlate different password practices with each other to determine which choices are made together in a general desire for higher security.

RQ4 How do sites' security requirements affect their implementation choices? The direct implications for both users and websites of password compromise vary greatly across sites. Users of sites that store payment information have a natural interest in the site keeping these details secure, as do the sites themselves, which may face legal consequences if this data is stolen. Users also have a keen interest in preventing compromise at sites storing personal information, such as email and social networks, while the sites themselves may have less economic incentive here. Still, to the extent that consumers may choose to join based on perceived security, all sites may have a general interest security. We can correlate different features offered by sites with implementation choices to assess how sites' functionality is reflected by their security policy.

RQ5 Why do websites choose to collect passwords? It is not always obvious why individual sites choose to implement a password collection system in the first place. While it may be mandatory for certain web services (such as webmail providers), e-commerce and news websites aiming at customisation only can be built without passwords by relying on client-side personalisation.

3.2 Selection of sites

We aimed to collect a large enough sample of sites to draw statistically valid conclusions and cover the major use cases for password authentication. Our first step was to explore each of the websites ranked in the top 100 worldwide in traffic by Alexa [6], a leading firm which compiles traffic data for the Internet. This initial survey led us to identify the following broad classifications of password-collecting sites:

- **Identity sites** Identity-providing sites enable users to create an online identity, and use passwords to restrict the ability to act on behalf of this identity. The largest examples are webmail, social networking, and blogging services, but there are many other types, such as gaming sites, forum-providers like the Internet Movie Database, or collaborative projects like the online encyclopedia Wikipedia. The key element of identity providers is that users are able to interact with other users with a persistent identity. While the importance of identities created varies, most users will have a vested interest in protecting the security of their account as they can build up a reputation over time. In many cases, such as webmail, users frequently entrust sensitive personal data to the site as well, and may even entrust the security of other accounts through password reset mechanisms.

Business models vary greatly amongst identity sites, but they are often based on advertisement or selling premium accounts.

- **E-commerce sites** E-commerce sites' primary purpose is to sell goods to their users. Many offer users the ability to create accounts to track their orders, save payment or shipping data, or store shopping preferences. There is often a direct possibility of fraud if accounts are compromised and payment details are stored, so e-commerce providers have an incentive to protect user accounts. In most cases users have no means of interacting with others on e-commerce sites. The identity they create only has a relationship with the merchant.

- **Content sites** Many sites collect password-protected accounts solely to allow users to customise the contents of the site. The predominant type of site in this category is news websites, typically online versions of traditional print media. News sites may provide a limited ability to interact with others by commenting on news stories (though in the vast majority of cases this can be done with no account), but the primary purpose of an account is to select which news a user is most interested in and receive email alerts of stories relating to their interests. This category is not exclusive to news. For example, Ask.com is a search engine which employs user accounts to customise search results, and Kayak.com is a travel-search website which uses accounts to save user preferences on travel. Neither site allows users to interact with other users, or sells any products of its own, so we would classify them as content sites.

These categories are neither completely exclusive nor exhaustive of all purposes for password collection. A large web conglomerate like Google could be argued to be in all three categories as it implements webmail, a news aggregation service, and the Google Checkout shopping service. We consider any site which allows significant user interaction with long-term identities (such as eBay) as an identity site, even if it also has merchant facilities. In practice we found only a few borderline cases, which were classified by hand, we also noted the actual features offered for additional analysis (see § 4.1).

Given these three broad categories of interest, we attempted to capture the largest sites within each category and also reflect the depth of the market by a random sample of medium-traffic sites. To accomplish this, we first attempted to find the 25 largest sites in each category. We included all 26 identity sites which were found in the general Alexa global top 100 list and added the top 25 sites from Alexa's specialised Shopping and News rankings. We then added a random sample of lower-tier sites. For the e-commerce and content categories, we took a random sample of 25 sites from the Alexa top 500 rankings in the Shopping and News categories. Collecting a random sample of identity sites is harder, as this is not a single category indexed by Alexa. We thus collected a random sample split between webmail and social networks, collecting 12 random sites from a Yahoo! directory listing free webmail providers,[2] and 12 random online social networks from a listing maintained on Wikipedia.[3] [4]

We note there were several limitations to our study which prevented some sites from being included. We limited ourselves only to sites offering free accounts due to our desire to collect a broad sample. This prevented us from signing up for members-only sites such as Netflix. However, we found that many sites we did sample offered premium accounts as well as free accounts using the same password system.

We also had to exclude banking websites, one of the most important applications of password authentication, due to the inherent difficulty of creating banking accounts, though we did note which sites we signed up for offered the option of storing payment details. Websites not offered in English had to be excluded for technical reasons, as were pornographic sites. Our total sample was 150 sites, a complete list of which, including categorisation, can be found in Table 9 in § A.1.

[2] `http://dir.yahoo.com/business_and_economy/business_to_business/communications_and_networking/internet_and_world_wide_web/email_providers/free_email/`

[3] `http://en.wikipedia.org/wiki/List_of_social_networking_websites`

[4] The sub-sample of identity sites is therefore split 26/24 between upper-tier and lower-tier sites due to sampling constraints, whereas the splitting is 25/25 for both e-commerce and content sites.

3.3 Evaluation process

Each site was evaluated in a scripted process. A number of automated tools were used to speed up evaluation and data recording, but there was a human in the loop at all times. The precise technical set up used to interact with sites is described in § A.2. The evaluation script proceeded as follows:

1. **Enrolment**—All website signup was done with an identical set of bogus user data. Each site requesting an email was given a unique mailing address, so that all subsequent marketing email could be traced back to its source. All password advice given to the user during the signup process was recorded, along with any data required to create the account. For initial signup, a relatively strong password '`pps2010!`' was provided, which was accepted at all sites.

2. **Login/Logout**—After initial signup, and email verification if required, the account was logged out of and logged into again. During normal login, the TLS details were observed if present,[5] along with data submitted to the server, and if a persistent login cookie was saved in the browser. After logging in, site details were examined such as what services could be accessed by the account, in particular if payment details were stored or premium account upgrades were possible.

3. **Password update**—After re-logging in, the password was updated through each site's "change password" interface. At this point, password requirements were tested by seeing which passwords would be acceptable.[6] First, length requirements were tested by attempting to enrol a short password. If the one-character password '`p`' was rejected, then concatenations '$\texttt{p}^n$' were tried until a minimum length n was established. Upper bounds were not tested, but were recorded if they were stated. Once length boundaries were tested, the passwords '`1234...`' and '`password`' were attempted. If these were acceptable (which occurred in the vast majority of cases), then it was assumed the site was placing no restrictions other than length restrictions. When these were rejected, more complicated passwords were tried to establish rules requiring letters, numbers, symbols, or non-dictionary requirements. Finally, a repeated password was tried to test if history was kept of previously used passwords. After update, the email account was checked to see if notification was sent of the password update.

4. **Password reset**—After updating the password, the account was logged off again. First, it was tested if feedback was provided to distinguish between a valid username and invalid password, or invalid username. Then, the password reset protocol was initiated. In most cases, this consisted of an email being sent with either the original password, a new temporary password, or a reset link. These were then used to log in to the account a final time to see if a password update was forced after completion of the protocol.

5. **Password probing**—Finally, we performed an automated brute-force attack on the newly create account, generating one login attempt per second with the correct username and random passwords. We ran the script with a limit of 100 attempts per site, stopping early if the site imposed a timeout, ban, or required CAPTCHA solving for further guesses.

It is important to note several limitations to this evaluation process. This was also a concern for past empirical studies of websites [35, 21, 48, 69], however sacrificing some depth of study is necessary to obtain sufficient breadth. Namely, we restricted ourselves to a black-box evaluation of the publicly observable facets of the sites' password implementation. There are many aspects to running a secure website which are impossible to assess from the outside, such as systems and network security in the sites' data centres. We made no attempt to study the security of servers against attack, say, by port-scanning them for known vulnerabilities. We also were unable to audit server practices such as storing large databases in cleartext. It was obvious in some cases, such as sites which send original passwords in email as a reset mechanism, but for most sites it was not observed but impossible to rule out.

[5] TLS data was recorded for all steps in which a password was entered.

[6] We chose to test allowable passwords in the update interface rather than enrolment interface for consistency, because enrolment is a one-time process whereas password update is repeatable.

Feature	Identity	E-Commerce	Content	Total
News displayed	15	0	49	**64**
Products for sale	4	50	1	**55**
Payment details stored	7	30	2	**39**
Social networking	28	1	2	**31**
Premium accounts available	17	3	8	**28**
Email accounts provided	17	0	2	**19**
Discussion forums	16	1	2	**19**

Table 1: Features offered by sites surveyed, sorted by overall prevalence. $N = 50$ for each category.

3.4 Supplemental data

In addition to our observed data, we used market research data provided by Alexa [6] to compare sites. This data includes traffic statistics for each site, as well as demographic data such as the site's popularity amongst different age groups, education levels, and genders. We correlate this data with our collected data on password implementations in § 5.

4 Data collected

Note on statistical significance

Except where noted otherwise, since our data is almost entirely categorical, we use use a G-test with one degree of freedom to test for statistical significance of correlation between two categorical labels applying together by placing the outcomes into a 2×2 contingency table [77]. We use a G-test with two degrees of freedom to evaluate a property's occurrence differing between three separate categories. Fisher's exact test is used for 2×2 contingency tables with entries so low as they make the use of the G-test inaccurate. We use the term "significant" to indicate a p-value less than 0.05, "strongly significant" to indicate a p-value less than 0.01, and "very strongly significant" to indicate a p-value less than 0.001. A p-value of less than 0.2 indicates an observed phenomenon approaches significance. Unless otherwise noted, two-tailed tests are used.

4.1 Site features

Because our main categories are broad, we collected orthogonal data on extra features offered by each site. The prevalence of these features by category is displayed in Table 1. The distribution of features demonstrates that our categories do correspond to meaningful differences—all of the features recorded differed between the three groups with very strong significance. Some cases of feature overlap were not surprising, such as the large number of identity sites with integrated news features. There are also a few outliers which highlight how large websites can take on unusual features: two news sites offered free webmail accounts (Times of India and Canada.com), two news sites offered social networking profiles (USA Today and The Lincoln Journal Star), the scrapbook supplies merchant Two Peas in a Bucket implemented social networking features and discussion forums, and the Bill O'Reilly Online news blog implemented its own apparel store. Of particular relevance for our analysis are sites which offer premium accounts and sites which store payment details, as both of these have stronger incentives to protect user accounts as password compromise enables fraud against the site.

Data	Identity	E-Commerce	Content	Total
Email address	38	50	49	**137**
Email updates offered	21	42	47	**110**
Postcode	15	30	34	**79**
Username	35	5	29	**69**
Mailing address	5	19	8	**32**
Phone number	5	20	7	**32**
Marketing data	4	6	13	**23**
CAPTCHA	29	3	11	**43**

Table 2: Personal data required for enrolment. $N = 50$ for each category.

Figure 1: Required collection of marketing data for registering with The New York Times.

4.2 Enrolment requirements

4.2.1 Personal data collected

Most sites required some personal information in order to create an account, the prevalence of various data items is displayed in Table 2. Many sites required additional personal information, ranging from address or telephone information to explicit marketing data (household income, job type, etc.), a typical example is shown in Figure 1. Content sites were significantly more likely to collect such data. They were also most likely to offer email updates, though this was common in all segments.

4.2.2 Email verification

After enrolment was completed, many sites immediately sent an email to the account provided. 65 sites required following a link in the email to complete the registration process. The breakdown of sites requiring email verification is striking—29 identity sites, 1 e-commerce site, and 35 content sites. Some identity sites made it clear they were using email verification to prevent fake account creation, particularly at sites providing free email accounts which were worried about spam. Newspaper sites have no security reason to verify email addresses, the fact that they did so indicates that gathering email addresses for marketing may be a primary purpose of their account systems.

Sending the user's cleartext password in a welcome email was observed at 16 sites. This dubious security practice occurred significantly more often at content sites (9 occurrences).

Advice	Identity	E-Commerce	Content	Total
Use digits	9	6	3	**18**
Use symbols	9	2	3	**14**
Graphical strength indicator	9	0	2	**11**
Difficult to guess	5	2	2	**9**
Not a dictionary word	6	0	2	**8**
Change regularly	4	0	1	**5**
Any	18	8	7	**33**

Table 3: Password advice given by sites surveyed. $N = 50$ for each category.

Figure 2: Password advice and a graphical strength indicator deployed on Microsoft Live.

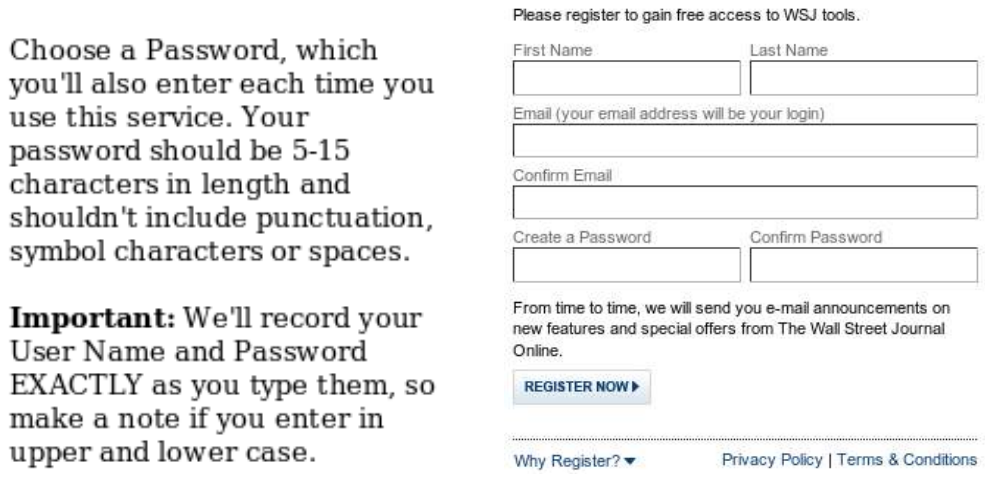
4.3 Password registration

4.3.1 Advice given

Given evidence that users will pick better passwords when given advice [13, 96, 118], it is surprising that 117 sites (78%) gave users no advice whatsoever on how to choose a password, as shown in Table 3. Identity sites were more likely to give advice with strong significance. Specific advice to include numbers or symbols was more common than vague recommendations to pick hard-to-guess passwords, though a few sites offered interesting advice, like Twitter's recommendation to "be tricky!". One site, Costco, provided a sample password '`RUGT_7`' which it described as strong.[7] Despite usability research suggesting their utility [29], graphical password strength indicators were uncommon (an example is shown in Figure 2). The long-criticised advice to change passwords regularly [13] was the rarest given, being mentioned on only 5 sites (3%). Only 5 sites give users the ability to register a "password hint" to be displayed in the clear to prompt them of their password. Though this feature has been deployed in Microsoft Windows for user accounts and has been recommended as a way of increasing password strength [96], it seems uncommon on the web.

The lack of password advice goes against both academic research suggesting its utility and US government FIPS standards on password collection [2, 26]. It is possible that sites believe that there is little security value in steering users towards choosing better passwords [39], but it seems most plausible that the lack of password advice serves to minimise the length of the signup form and increase enrolment rates. An instructive example is shown in Figure 3, comparing the advice given at The Wall Street Journal today to that given 15 years ago when the site first launched. The entire modern registration form, implemented as a pop-up frame in JavaScript, takes up less screen space than just the textual description of passwords that was provided in 1996. The vast majority of modern sites surveyed gave no indication of what a password is beyond the word "password" and a password-input box, expecting the

[7] While this may be a reasonably strong password, it is statically generated, which means all users receive the same example password, which may be a good value for attackers to guess.



(a) 1996 (b) 2010

Figure 3: Password advice given at The Wall Street Journal in 1996 (left) and 2010 (right). Note the advice against special characters in the early version.

user to already understand the concept and have the ability to choose an appropriately secure password.[8]

The requirement to enter one's password twice upon enrolment was very common, with 132 sites (88%) doing so. Interestingly, this requirement was significantly less common at identity sites, representing 12 of the 22 sites which only required one entry of the password. It appears that sites with very minimal enrolment forms, particularly implemented in popup JavaScript frames directly from the site's home page, dropped this field to save space in the enrolment form.

4.3.2 Password requirements

The most common restriction placed on acceptable passwords was length, with 123 sites (82%) imposing a minimum length. Among sites with a minimum length, 6 characters was by far the most common requirement, occurring in 78 sites (52%), 4 and 5 character limits were next with 21 and 15 sites (14% and 10%), respectively. This is consistent with Furnell's finding with a smaller sample size [42] and may in fact date back to the FIPS publication of the 1980's recommending six-character passwords for security [2]. Only 4 sites had a requirement of more than six characters in a password. The distribution of minimum password lengths is plotted in Figure 4. There was very little difference in minimum length requirements between identity, e-commerce, and content sites. Sites offering premium accounts and sites collecting payment details were significantly more likely to have minimum password requirements.

Aside from length, very few sites imposed any password restrictions. Only 14 sites (9%) implemented basic dictionary checking which prevented '`password`' from being accepted, despite empirical studies showing that many users will choose one of a few very common passwords if given no restrictions [80, 66, 101, 110]. 11 of the sites checking for dictionary words were identity sites, 5 offered premium accounts and 6 to store payment details, all of which were significant correlations.

More complex requirements were even less common, with only 7 sites requiring a digit placed in the password, and 2 sites requiring non-alphanumeric symbol characters. 2 sites placed a prohibition on reusing a password which a user had previously registered with the system, including eBay, which was the only site to give users an option to have forced password updates at regular intervals. This suggests that the oft-criticised requirement of regularly changing passwords [17] is more popular for system accounts and is not common on the web.

[8] A notable exception was Hushmail, a webmail provider billing itself as the "most secure web-based free email service in the world." Hushmail uses the term "passphrase" and gives users a paragraph of advice, including a recommendation to use a secure-password generating application like Diceware [7].

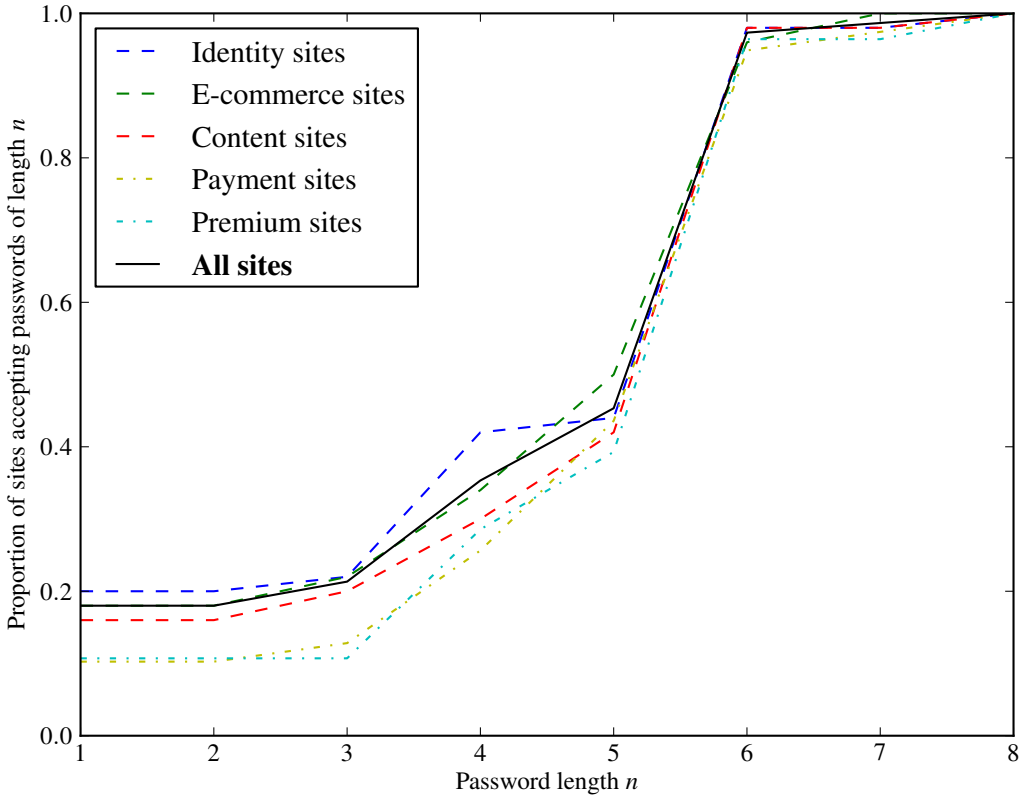


Figure 4: Cumulative istribution of minimum password lengths at different site categories

One site, Fertility Friend, did not allow users to choose their own passwords initially, instead piggybacking on the email verification step to send new users a randomly generated password which was composed of four random text characters. Users were allowed to change their password after logging in once with the randomly generated one. It has been long known that system-generated passwords will be stronger than user-selected passwords [80], but evidently the usability of this is too low for commercial use. The security of emailing a random password is also questionable, as it will remain in cleartext in the user's inbox if they don't change it.

4.4 Login

4.4.1 Identification

The vast majority of sites (130, 87%) allowed users to identify themselves during login using their email address in lieu of a site-specific username. Only 69 sites (46%) allowed for the creation of site-specific non-email usernames, with 6 not even accepting these for identification purposes. This was particularly true at e-commerce sites, only 5 of which offered usernames. This trend of email addresses replacing usernames seems an acknowledgement by sites that users now maintain too many accounts to remember a separate username at each one, despite the fact that passwords are still collected. It is perhaps more surprising that such a large number of content sites still require usernames, many of which didn't even have commenting capabilities for which a username would be useful as a pseudonym. Of the 20 sites not accepting email for identification, a very strongly significant majority of 13 were identity sites. Most of these were email providers themselves which refused registration of a secondary email address. 61 sites allowed users to log in with either their email address or username.

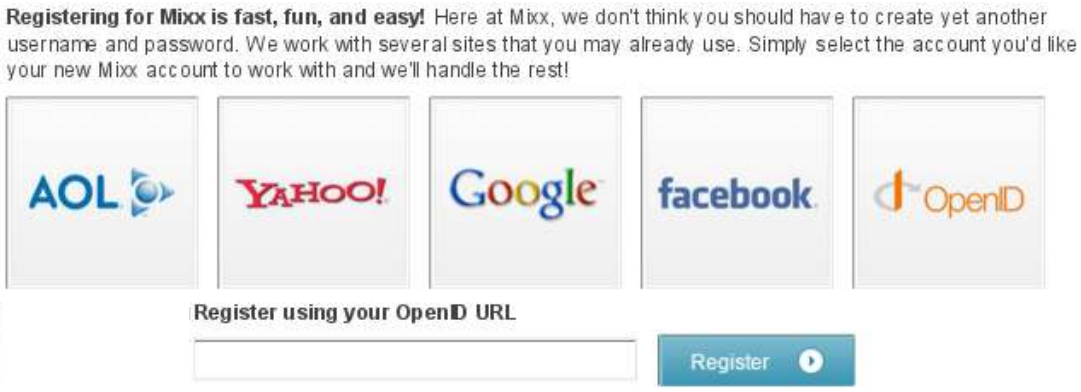


Figure 5: Recommendation to use OpenID, with an enumeration of popular providers, Mixx.

4.4.2 Password submission

All sites surveyed used an HTML `input=password` form element to capture passwords during regular login. Two sites enabled password entry into an un-starred `input=text` element at any stage (the passsword update interfaces at Art Beads and DVDEmpire). One site, Gamespot, submitted the password both in the POST form data and in the URL target of the POST, a bizarre design which leaves the password in browser history. No sites utilised the built-in HTTP authentication mechanism, which enables password entry in a browser-supplied popup window or a client certificate [40]. TLS deployment to protect password submission was mixed, as is discussed in § 4.10.

Only three sites took the security step of hashing the user's password in JavaScript running in the browser during normal login, preventing the server from ever receiving the user's cleartext password. Of these, Microsoft Live still collected cleartext passwords during enrolment, and Bodybuilding.com did so apparently by accident, hashing the first entry of the password prior to submission but still submitting the re-entry of the password in cleartext. Only Ask.com, a customisable search engine, never submitted cleartext user passwords to its own servers, providing a firm guarantee to users that the server was not storing the user's password in a cleartext database. Interestingly, two sites (Hushmail and Swiss Mail) claimed never to submit unencrypted passwords to their servers, but inspection of the actually generated HTTP POST packets revealed this was not the case.

4.5 Federated identity

Federated identity (or single sign-on) systems can obviate the need for users to store passwords at many sites by allowing a single server which knows their password to verify their identity to others as they sign up for new services (§ 2.5). Support for these systems, in the form of allowing users to create an account with registering a password, was extremely low. Windows Live ID, formerly known as Microsoft Passport [10], is one of the oldest such systems, though it appears that privacy and security concerns have completely killed it [86]. It was not accepted by any of the sites we surveyed.

OpenID [92], a decentralised standard which has received much attention, was rarely accepted as a primary means of authentication. It was accepted at only 4 sites, all in the identity segment. A commendable example was Mixx, which gave prominent billing to OpenID and recommended against users registering a password, as seen in Figure 5. Facebook additionally gave users the option of logging in with an OpenID, but required first registering a password with Facebook which largely undermines the purpose of the protocol. More sites (6) were OpenID providers, with 5 being providers only and just one site (LiveJournal) acting as both a provider and an relying party. This is consistent with previous analysis that large sites are more willing to provide OpenID authentication than to rely on it themselves [15].

The interfaces provided by sites which do act as OpenID providers also supported the notion that mainstream deployment is not the ultimate goal. Google provides OpenID, but this was undocumented

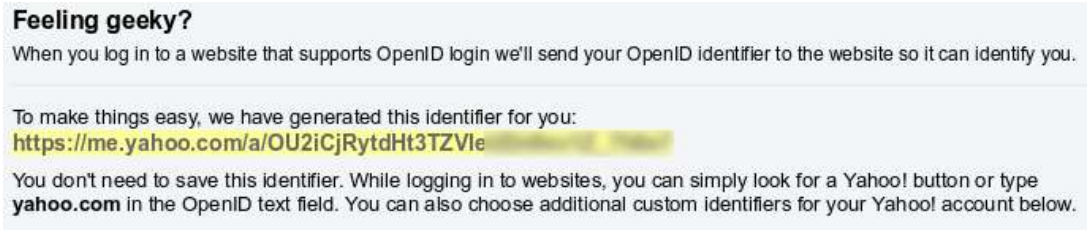


Figure 6: Obfuscated description of OpenID at Yahoo!.

anywhere on the site, and required searching for instructions on an external site to figure out the format of Google OpenID URLs. Yahoo! gave users a nearly impossible OpenID URL to remember, including a 32-character base-64 encoded string as seen in Figure 6, instead of a human-memorable URL like `http://me.yahoo.com/user_id` as is suggested. Yahoo also described OpenID as "geeky."

Facebook Connect, a younger, proprietary protocol similar to OpenID, was more widely deployed, with 10 sites accepting Facebook Connect authentication instead of passwords (though this difference was not significant). Also in contrast to OpenID, three sites outside the identity segment (all content sites) were willing to accept Facebook Connect authentication.

4.6 Password update

Most sites implemented an interface to update one's password to a new value. Only 3 sites didn't offer password update, instead using the password recovery mechanism to allow users to update their passwords. The threat to password update mechanisms is that an adversary can hijack a user's session or temporarily use a logged-in browser in to change a victim's password to a known value, particularly relevant since every single site in our survey gave users the option of storing persistent login cookies.[9]

Most sites (111, 74%) required re-entering the existing password before it can be changed as a defense against this, with identity sites being significantly more likely to do so than the others. 21 sites took the step of sending an email notifying the account holder that their password has been updated. E-commerce sites were significantly more likely to take this approach. Bodybuilding.com offered users an amusing message after password update: "Congratulations, your ex can no long access your account!"

4.7 Password recovery

All sites except two implemented a means of password recovery in case a password is forgotten. Hushmail, designed for security, warned users from the start that passwords are not stored and all email is stored encrypted, so there is no recourse for a lost password. The Fort Worth Star-Telegram took the refreshing step of simply telling users to create a new account if they have forgotten their password.

4.7.1 Email-based recovery

The vast majority of sites (138, 92%) offer email-based password recovery, sending reset instructions to an email address supplied during enrolment, with 18 (12%) of them additionally requiring the user to answer personal knowledge questions, as displayed in Table 4. In general, sending out time-limited password reset emails is considered secure, though there are several technical pitfalls [44]. The contents of the recovery email came in three main forms: sending the user's original password, sending a new randomly-generated password, or sending a one-time link to access a password update interface (also displayed in Table 4).

[9] Research has also shown that many sites create their authentication cookies insecurely [41], though we did not attempt to evaluate this in our survey.

Recovery Mechanism	Identity	E-Commerce	Content	Total
Email only	32	42	46	**120**
Email plus personal knowledge	11	4	3	**18**
Personal knowledge only	5	2	1	**8**
None available	2	2	0	**4**
Email contents				
Original password (cleartext)	5	14	17	**36**
Temporary password	11	15	12	**38**
Reset link	29	18	20	**67**

Table 4: Password recovery mechanism frequency. $N = 50$ for each category.

Sending the original password in cleartext (Figure 7) is certainly a mistake from a security perspective; this practice occurred less often with identity sites with strong significance. The best solution, sending a time-limited reset link (Figure 9), was implemented about half of the time, with identity sites being very significantly more likely to implement this. Sending a new randomly-generated password (Figure 8) is also undesirable, because users may elect not to change it and this then presents the same problem, although many sites require password update immediately after logging in. 70 sites overall require reset after the recovery process is complete, with identity sites being more likely to do so with strong significance. The online merchants Gap and BestBuy offered a nice security feature, automatically deleting all of a user's stored payment info after password reset.

4.7.2 Personal knowledge questions

A smaller number of sites (26, 17%) allow password reset by answering personal knowledge questions. Research has shown this to be a risky form of backup authentication, as personal knowledge questions are easy to lookup online [90] or in public records [49], and are easy to guess for acquaintances [97] or using known statistics of likely answers [20]. In the specific case of backup authentication online, email authentication has been argued to be more secure than personal knowledge questions if implemented properly [44, 64]. Still, webmail providers must be able to cope with users who have no secondary email account, hence identity sites were very significantly more likely to allow this feature, with 16 identity sites (32%) collecting personal knowledge questions. 7 allowed the user to choose email resets if they had a backup email address. The fact that 6 e-commerce sites and 4 content sites collected personal knowledge questions is highly dubious, though 4 of these sites require both personal knowledge questions and email. One site, Mail.com, attempted to do this but allowed the user to specify which email address to send a temporary password to, eliminating the security of the combined approach.

There is also a concern for usability with the proliferation of sites, as users have a very small number of personal knowledge questions to draw on. Most site implemented the same small group of personal knowledge questions, with 24 of 26 sites using either mother's maiden name or pet's name as a backup question. 4 sites allowed users to choose their own challenge questions, despite research showing that users will choose very insecure questions when given the opportunity [63].

4.7.3 Other recovery methods

A few sites innovated on the basic model. Some did not explicitly collect personal knowledge questions. Phillyburbs.com, for example, required the user to enter his or her postcode prior to getting a reset email. Google offers users the opportunity to have password reset instructions sent to their mobile phone via SMS. No sites were observed implementing "social" backup authentication as has been proposed in the

```
Below is your login information which you requested from our site.

When prompted to login or to modify your account, enter the
following exactly as shown:

Email address:  <email>
Password:  <clear-text password>

Thank you for visiting chicagotribune.com.
```

Figure 7: Password reset via sending original cleartext password, The Chicago Tribune.

```
Hello, <username>:

Thanks for using your Ticketmaster account.

This is a temporary password:  <temporary-password>
Use this temporary password to login and reset your password
again.

We hope you enjoy using your account!

Thanks,
The Ticketmaster Team
```

Figure 8: Password reset via temporary password, TicketMaster.

```
Hi <username>,

Someone requested that your Last.fm password be reset.  If this
wasn't you, there's nothing to worry about - simply ignore this
email and nothing will change.

If you DID ask to reset the password on your Last.fm account, just
click here to make it happen:
http://www.last.fm/?id=<userid>&key=<authentication-token>

Best Regards,
The Last.fm Team
```

Figure 9: Password reset via temporary reset link, Last.fm.

limit	countermeasure				
	CAPTCHA	timeout	reset	unknown	total
3	(3, 0, 0, **3**)	–	–	–	(3, 0, 0, **3**)
4	(1, 0, 0, **1**)	(0, 1, 0, **1**)	–	–	(1, 1, 0, **2**)
5	(2, 1, 1, **4**)	(0, 0, 2, **2**)	(1, 1, 1, **3**)	–	(3, 2, 4, **9**)
6	(2, 0, 0, **2**)	–	(0, 2, 0, **2**)	–	(2, 2, 0, **4**)
7	–	(1, 0, 0, **1**)	–	–	(1, 0, 0, **1**)
10	(2, 0, 0, **2**)	–	–	–	(2, 0, 0, **2**)
15	–	(1, 0, 0, **1**)	–	–	(1, 0, 0, **1**)
20	(0, 1, 0, **1**)	–	–	–	(0, 1, 0, **1**)
25	(1, 0, 0, **1**)	–	–	–	(1, 0, 0, **1**)
> 100	–	–	–	(37, 43, 46, **126**)	(37, 43, 46, **126**)
total	(11,2,1,**14**)	(2,1,2,**5**)	(1,3,1,**5**)	(37, 43, 46, **126**)	–

Table 5: Password guessing limits and countermeasures. For each combination of cutoff and response mechanism, the number of implementing sites in the identity, e-commerce, and content segments is shown in the tuple $(i, e, c, i + e + c)$. The tuple $(0, 0, 0, 0)$ is represented by '–'. For each category, the total number of sites is $N = 50$.

literature [98], nor did any sites implement an "adaptive" challenge question scheme, querying the user based on their past interaction with the site [16].

4.8 Rate limiting for password guessing

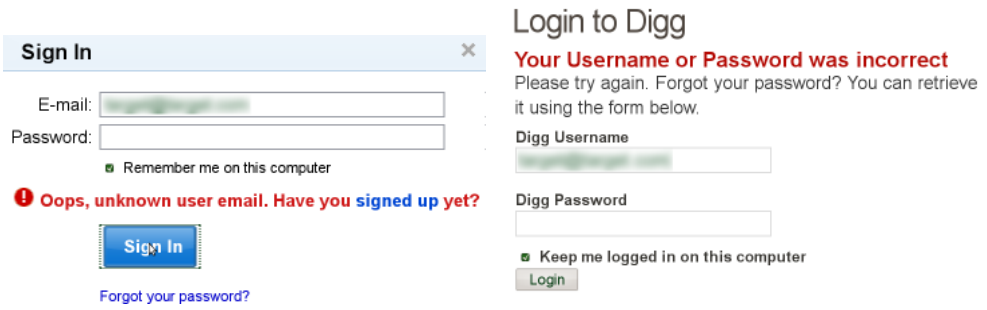
A basic assumption made in most literature on password authentication is that guessing attacks against websites are 'online,' meaning that it is possible to limit the rate at which adversaries can guess passwords.[10] We measured the number of attempts allowed using an automated script to guess randomly-generated passwords at a rate of one per second. We assume the adversary has a list of enrolled usernames or email addresses to guess passwords for. As discussed in § 4.9 this is almost always obtainable from websites themselves by probing for user existence prior to conducting a guessing attack. Thus, we tested for valid accounts only. A small number of sites re-directed the user to the password reset screen after a cutoff number of guesses, but allowed further guessing at the log-in screen. This feature has no security value as an attacker will simply ignore the re-direct, as we programmed our script to do.

The vast majority of sites surveyed (126) allowed our script to guess 100 passwords with no restriction, at which point we stopped it and were able to successfully log in.[11] It is impossible to say for certain that a site implements no guessing cutoff, but we considered it unlikely that any site would have programmed a cutoff greater than 100. Thus, most sites do not take seem to take advantage of their theoretical ability to limit guessing attacks, including a number of surprisingly high profile sites such as Amazon, eBay, and WordPress.

Of the sites which did implement restrictions, there was a wide variety both of cutoff thresholds and measures to limit further guessing, as seen in Table 5. Most sites implementing a cutoff allowed 3 to 6 guesses without restriction, with 5 the most common limit. This goes against the "three strikes and you're out" conventional wisdom, although it is shorter than the 10 guess limit recommended in an academic usability study [23]. A few sites allowed higher limits, with Yahoo!'s limit of 25 being the highest observed. Identity sites were significantly more likely to restrict guessing.

[10] An alternate suggestion to prevent brute-force attacks is a large number of honeypot accounts [55], we were unable to test if this strategy has been deployed due to our black-box approach.

[11] It would be possible for a site to allow unlimited guesses, but simply reject all guesses after a certain cutoff, however, all sites tested allowed us to log in with the correct password after our guessing was finished.



(a) Ask Jeeves (unprotected) (b) Digg (protected)

Figure 10: Probing for user membership at login

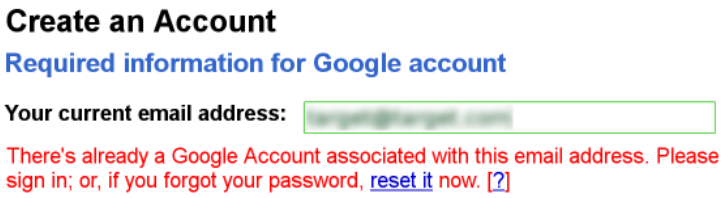


Figure 11: Probing for user membership during enrolment, Google.

The most common countermeasure was requiring that a CAPTCHA be solved with each additional password guess, implemented by a strongly significant majority of 14 of the 24 sites (58%) with a guessing limit. Other approaches included triggering an automatic password reset (5 sites), or imposing a timeout before further guesses could be made (5 sites).

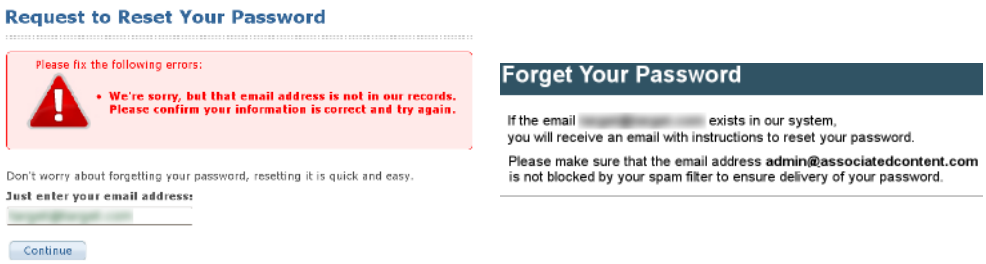
4.9 Prevention of user probing

It has been advised for security that sites make it difficult for attackers to determine the list of enrolled usernames to mitigate trawling or "horizontal" attackers attempting to guess a few likely passwords for a large number of users [39, 20]. Furthermore, because users are likely to re-use both email addresses and usernames for accounts on different sites, it is important not to reveal either if they are accepted for identification at login, as they can then be used in a password guessing attack. If an attacker wishes to comprise Alice's password at site X, he may check if Alice is registered at sites Y, W, and Z, providing multiple points of attack if Alice has re-used her password.

Thus,it is undesirable for sites to expose an interface for attackers to automatically probe for the existence of either a username or email if they are also accepted as identification at login (different sites choose only one, or both, of these options, as described in § 4.4.2). There are multiple interfaces which can be used for probing, which we discuss in turn: login, enrolment, and reset.

4.9.1 Login

The login screen can be used for probing if different error messages are given for an invalid identifier and an invalid password, as seen in Figure 10(a). Most sites avoided this problem, with 122 (81%) giving a generic error for either incorrect identifier or incorrect password, as seen in Figure 10(b), with no significant difference between site categories. The general resistance to this form of probing establishes either that most sites do consider it a threat, or possibly that implementing a generic error is the simplest possible choice.



(a) Zappos! (unprotected) (b) AssociatedContent (protected)

Figure 12: Probing for user membership during password reset

4.9.2 Enrolment

An attacker can also probe for the existence of identifiers by attempting to register new accounts with a target identifier. In some sense, this is a fundamental problem because sites offering free accounts must let a user know if their enrolment succeeded. However, this probing point can be limited by checking a CAPTCHA before acknowledging if an identifier is available for a new account. Most sites instead provided instant feedback about the availabilty of identifiers, as seen in Figure 11. Only 7 sites (5%) protected against user probing at enrolment by reporting success only after a CAPTCHA, of which a significant majority (5) were identity sites.

4.9.3 Reset

Attackers may also probe the password reset interface by attempting to reset the password for a target identifier, as seen in Figure 12(a). This weakness has been noted before [22] and remains prevalent, with 130 sites (86%) revealing membership through password reset requests. This included the vast majority of content sites (48/50), more than the other segments with very strong significance. 10 sites (7%) implemented CAPTCHAs to prevent probing the reset interface and 10 gave messages to the effect that "If the username you entered is in our system, we will send an email with password recovery instructions," as seen in Figure 12(b).

Only 2 sites didn't allow unrestricted probing at any of these three points (IKEA and MySpace), while 22 sites (15%) didn't restrict probing at any point and 126 (84%) had a mixed approach. Of these, most protected login but not enrolment or reset, however, 6 sites didn't protect login but did protect at least one of the other two interfaces.

4.10 Encryption and authentication

Different sites implemented TLS to protect password submission in different ways, as displayed in Table 6. There are up to four separate forms for password entry: enrolment, login, password update, and password update after reset. 59 sites (39%) deployed TLS properly for all password entry forms, and 61 sites (41%) offered no TLS at all. TLS adoption is higher amongst e-commerce sites with very strong significance, which is not surprising given these sites' need to process payment information. Similarly, with very strong significance, content-only sites deploy TLS less frequently than either other category.

Perhaps the most interesting finding is the high rate of inconsistent deployments of TLS, protecting some password entry forms but not others. Sites typically protect the normal login form but forget to protect the enrolment, update, or recovery update interfaces, but two sites (The New York Times and CD Wow!) managed to protect all password entry mechanisms except for normal login. 25 different sites made such errors in TLS deployment, thus over 28% of sites which made any effort to deploy TLS have a broken implementation, including very high-profile sites such as Facebook, MySpace, Twitter, WordPress, and LiveJournal.

Deployment Level	Identity	E-Commerce	Content	Total
Full	10	39	10	**59**
Full/POST	3	1	1	**5**
Inconsistent	14	6	5	**25**
None	23	4	34	**61**

Table 6: TLS deployment. $N = 50$ for each category.

A small number of sites implemented TLS only for the POST action of the login form. This enables login directly from a public facing (non-TLS) version of the site, while still protecting passwords.[12] What is most interesting though is that of the 18 sites (12%) which implement POST-only TLS for login (12 of which are identity sites), a strongly significant majority (13) forgot to use TLS on other password submission forms on the site, with only 5 successfully implementing a mixed approach.

Only two sites enabled a user option for TLS submission of passwords (LiveJournal and Mail2World), which anecdotally suggests that user-optional TLS is an obsolete feature from the early days of inconsistent TLS support in browsers.

5 Analysis

Note on password security scores and policy tuples

In order to facilitate analysis, we defined a 10-point scoring system which measures the overall level of password security provided by each site. This score should not be interpreted as definitively meaning some sites are better or worse, only measuring the number of steps each site has taken to prevent password compromise. Some academic research suggests against preventing weak passwords [39], giving users too much password advice [13], or heavily restricting the number of login attempts allowed [23], but we consider all to indicate more security-conscious sites. Two sites with the same password score may still differ in their exact password protection measures. The complete scoring formula is provided in § A.3.

We further defined a condensed format for representing the major security policy decisions made by each site. This is a tuple of ten features representing the major dimensions of security policy each site has designed, which can be used to cluster sites by their policy choices. The complete definition is provided in § A.4.

5.1 User experience

Returning to **RQ1**, we see consistency in the basic password mechanics but large and in many cases unnecessary variations in measures to improve security.

Basic password entry is consistent

Nearly every site utilised the HTML `type=password` input field, and the word "password" itself. 87% of sites accepted an email address as the identifier for login, and every single site surveyed utilised persistent log-in cookies to prevent frequent password entry. 78% of sites provided no advice or guidance on what a password is, demonstrating that users are expected to have internalised the concept of web-based password login.

[12] This practice is not ideal, as a network attacker can re-write the webpage to direct password submission to his own server [35]. However, user studies show a very low rate of awareness when TLS indicators are turned off [99].

Security advice and requirements are inconsistent

Where they existed at all, both password requirements (§ 4.3.2) and password advice provided (§ 4.3.1) varied considerably from site to site even with specific industries. Sites choosing to give advice all provided unique strings describing how to choose a password; the only consistent advice given was none at all. We found only two sites which provided links to third-party sites providing advice on password selection. Other user-visible aspects of security, namely TLS deployment (§ 4.10), and password reset mechanisms (§ 4.7), also varied considerably even within market segments with similar security requirements. Since a lack of user understanding of password threats has been shown in surveys [13, 93, 45], this inconsistency is a major problem.

5.2 Security weaknesses

Returning to **RQ2**, we find that security is undermined by inconsistency and sloppy implementations.

Best practices are far from universal

Most aspects of password security best practice that the existing literature has agreed upon were found missing at a significant number of sites. Specifically, 57% of sites failed to use TLS to protect password transmission in all cases (§ 4.10), 29% emailed cleartext user passwords indicating that they are not hashed prior to storage (§ 4.2.2 and § 4.7.1), 83% allowed attackers unrestricted probing of user membership (§ 4.9) and 84% allowed unrestricted guessing of passwords (§ 4.8). These practices can be interpreted unambiguously as mistakes from a security perspective.

Many aspects of password implementation are not standardised

There are several elements of password security for which there is no generally agreed-upon practice, such as what measures should be taken to prevent weak passwords [39], how much password advice users should be given [13], the number of login attempts which should be allowed [23], or the implementation of password reset [64]. Amongst these, there is strong diversity in the wild. Password requirements beyond length were uncommon, but varied between requirements of numbers or symbols (§ 4.3.2). Of sites implementing a minimum password length, the majority (61%) chose 6 characters, but there was substantial variation beyond this (§ 4.3.2). The most popular limit for guessing attacks (§ 4.8) was 5 guesses (38%), but other sites were evenly split between less than 5 guesses, 6-10 guesses, or greater than 10 guesses (21% each). Reset links were the most popular mechanism for email-based password recovery (48%) but temporary passwords (27%) and cleartext passwords (25%) also were widely implemented (§ 4.7).

Many security policies are internally inconsistent

Beyond the differences between policies at different sites, many sites have implemented policies that are internally inconsistent, specifically with regards to TLS deployment where 29% of all sites implementing TLS forgot to apply it in all cases (§ 4.10), and preventing user probing where most sites protected some but not all interfaces (§ 4.9). This suggest more directly that many sites' policies may not have been intentional but may have simply been implementation decisions taken by programmers.

Source code does not appear to be widely shared

In addition to the large number of different approaches, we find indirect evidence based on qualitative inspection of HTML and JavaScript that sites are not re-using standard source code to implement password authentication, with individual site developers apparently re-implementing what could be standard functionality. We were able to corroborate this observation by parsing password registration and recovery emails and seeing little evidence of identical formats between sites. In all four cases where we did

see evidence of repeated implementations, we were able to directly find that the similar sites were again in fact owned by a common parent company. To the best of the authors' knowledge, there does not exist a widely-used open-source implementation of password authentication.[13]

Security policies vary far more than security requirements

Within the e-commerce and content segments, the majority of sites have very similar security requirements and thus could implement a standardised security policy for password collection, entry, update, and reset. We condensed each site's choices into a tuple of ten policy choices for passwords (§ A.4) to examine how similar sites were and found an incredible variety of policies. At this fairly coarse granularity, we identified 142 distinct policies in the sites we examined. Of sites sharing policies, one group of three newspaper sites (the Chicago Tribune, the Los Angeles Times, and the Orlando Sentinel) are indeed all owned by a common parent company (The Tribune Company). Searching for policies which are similar in most elements confirmed the variety of implemented policies. The majority (56%) of sites surveyed were not within a Hamming distance of 1 of any other sites in terms of policy, and a sizeable number (24%) were not within a distance of 2 from any other site. The policies of IKEA and LiveJournal stood out as notable outliers, neither being within a distance of 4 of any other policy. Not coincidentally, both received the highest possible scores in our evaluation. The low state of the art means that the best implementations are frequently pioneers, different from most other sites.

Clustering sites by their policies is instructive (details of our clustering approach are provided in § A.5). Using a maximum radius of 3 for each cluster produces 39 clusters of sites, as shown in Figure 13, The largest cluster contains 68 sites, 45% of all those studied, and represents a standardised middle ground of password implementation. While there is variation within this middle group, it is generally marked by TLS deployment, no resistance to password guessing or user probing, recovery via email-based reset, no password advice, no limitations on passwords except for a 6-character minimum, no support for federated log-in. The large cluster has a slightly above-average password score and is significantly biased towards e-commerce sites, which make up half its members. Beneath the large cluster are a number of medium-sized clusters of sites with well-below average scores of sizes 21, 11, 5, and 3 sites, each of which is significantly biased towards content sites. There are several small clusters of better-than average sites (of size 4, 3, 2, and 2). There are then 29 singleton clusters, of which the majority (69%) are above-average sites. Taken together, the singletons are significantly biased towards identity sites.

The overall picture indicates that at a coarse level, there is a large middle-ground dominated by e-commerce sites which provide a minimal password implementation (with several commonly overlooked holes), followed by several medium-sized groupings of similarly implemented low-security sites led by content sites, and identity sites offering more security and diversity in their password implementations.

[13] There do exist commercial implementations for sale such as `http://www.sentrylogin.com/sentry/index.asp`, `http://www.monster-submit.com/sentry/` and `http://www.authpro.com/faq.shtml`.

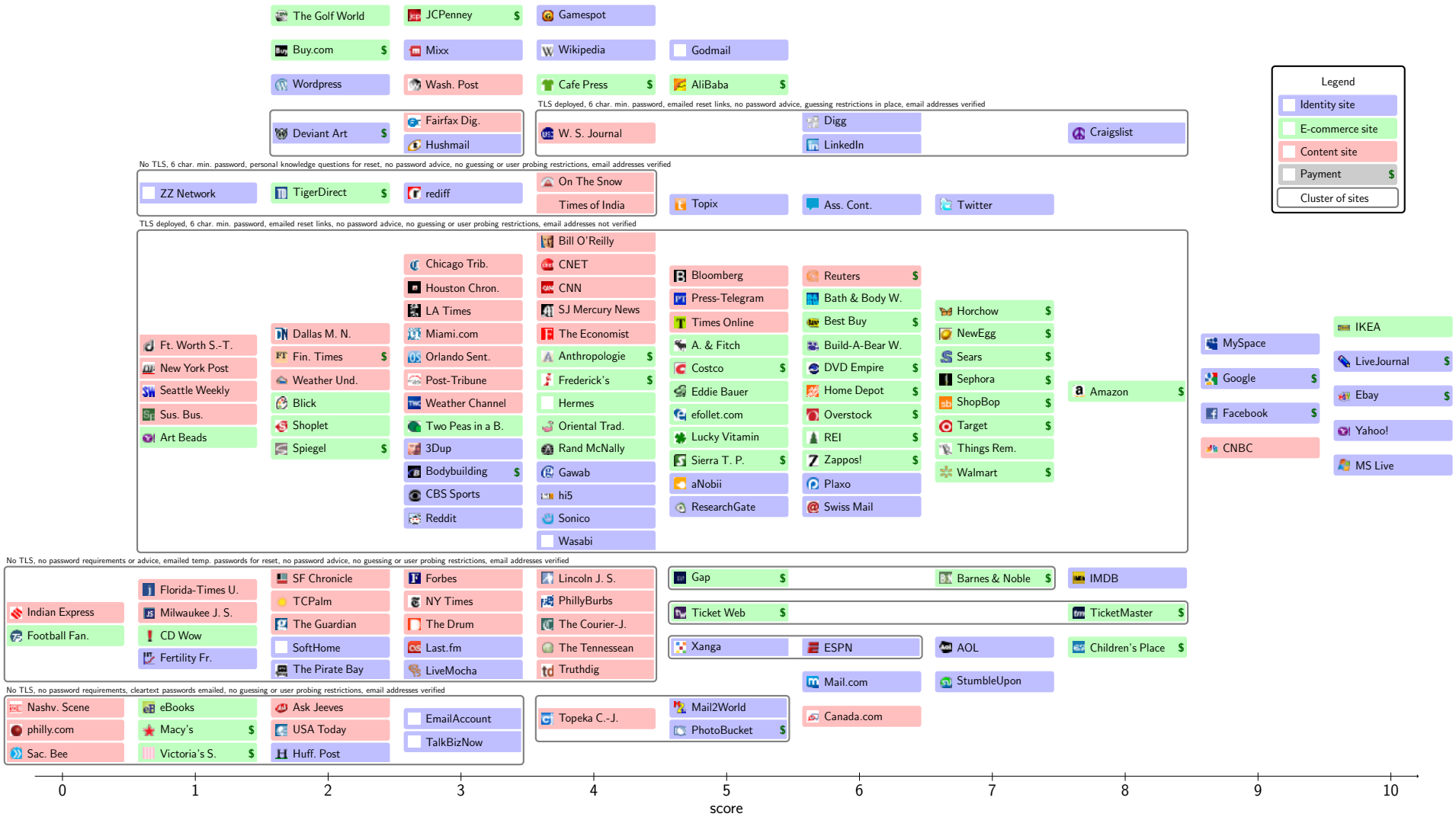


Figure 13: Clustering of surveyed websites by password security policy (§ A.4). Clustering was performed using the QT clustering algorithm with a maximum cluster radius of 3 (Hamming distance), details are provided in § A.5. The horizontal axis orders sites by their password score (§ A.3).

	Password score > median	TLS deployed correctly	Guessing attacks restricted	Minimum password length enforced	Dictionary words prohibited	Cleartext passwords mailed	Notification of password reset	Email verified on enrolment	CAPTCHA required on enrolment
Positive 3-mo. traffic change	↑↑	+	↑↑↑		↑	+	+		
Years online > 10		↑↑	↓↓	+				↓	↓
Load time < med.	↑	↑	↑		↑	−	↑	↓↓↓	
Traffic Rank > 25th %ile	↑↑↑	↑	+	+			↑↑	+	
Traffic Rank > med.	↑↑↑		↑↑	+	↑↑↑	↓	↑	+	+
Traffic Rank > 75th %ile	↑↑↑		↑↑↑	↑	↑↑↑	↓	+	↑↑↑	↑↑
Industry Traffic Rank > 25th %ile	↑↑↑	+	+	↑			↑	+	
Industry Traffic Rank > med.	↑↑↑	+	↑↑↑	↑↑↑	↑↑↑		↑↑		
Industry Traffic Rank > 75th %ile	↑↑↑	↑	↑↑	↑	↑↑	−	↑↑	+	
Page Views > 25th %ile	↑↑↑	↑↑					↑↑		
Page Views > med.	↑↑↑		↑↑	+	↑↑↑	↓	↑	+	+
Page Views > 75th %ile	↑↑↑		↑↑↑	+	↑↑↑	↓↓	↑	↑↑	↑↑↑

Table 7: Correlation between traffic statistics and observable indicators of password security, including our aggregate password score defined in § A.3. Statistical significance is indicated by (+/−) ($p \leq 0.2$), (↑/↓) ($p \leq 0.05$), (↑↑/↓↓) ($p \leq 0.01$), and (↑↑↑/↓↓↓) ($p \leq 0.001$).

5.3 Security performance and market position

Addressing **RQ3**, we analyse the relationship between a sites' numerical password score and general market data about the site published by Alexa [6], a general-purpose market-research firm. We correlate market statistics with our overall password score as well as two simpler baseline indicators of password security: TLS deployment (§ 4.10) as well as implementation of restrictions on password guessing (§ 4.8). In general, we find that password security increases with size, traffic, age, and engineering quality of a site; a summary of correlations found is shown in Table 7. This is consistent with the fact that implementing password security imposes non-trivial engineering overhead costs. Larger sites with more engineering resources are thus more easily able to invest in security.

Leading sites deploy better password security

More popular sites have better password security scores with strong significance. This correlation holds with very strong significance across a range of measures, using either Alexa's proprietary traffic ranking or raw page view numbers, and dividing sites against the median, 25th or 75th percentiles. The correlation is more sporadic for TLS deployment, as a number of leading sites, due partially to the frequency of errors in TLS deployment even at very high-traffic sites (§ 4.10). Correlation to guessing restrictions (§ 4.8) only exists for very-high traffic sites, as restrictions on guessing are relatively rare.

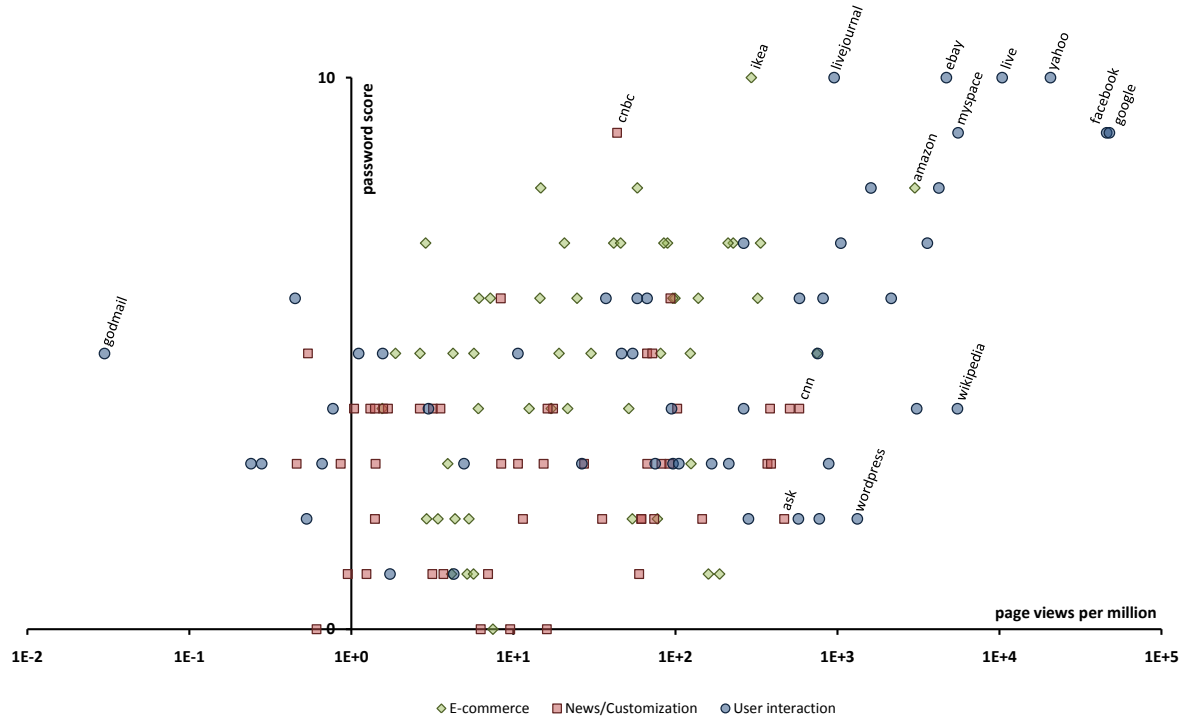


Figure 14: Password scores from 0 to 10 plotted against the raw Alexa page view numbers for each site on a logarithmic scale. Notable outliers are annotated.

Industry leaders have more sophisticated password practices

To control for industry-specific differences in popularity, we analyse a site's traffic rank relative to its own category as a rough assessment of market popularity relative to competing firms. The trend towards better password practices at more popular sites remains very highly significant.

In Figure 14, password scores are plotted against the raw number of page views each site attracts, labelled by site category. The overall trend towards better security at more popular sites within each industry is clearly visible, as are a number of outliers. Within each category we observe an upward-shaped lasso: some highly popular sites extend the lasso to the right, but highest scores are achieved by close followers to the market leader (e.g. IKEA instead of Amazon; Yahoo! instead of Google).

Sites with more sophisticated password practices grow ahead of the market

For most websites, traffic growth directly tracks business success. Sites whose traffic rank is increasing are typically increasing market share within their industry. We use the three-month variation in Alexa traffic rank to assess growth and distinguish between negative variation and stationary or strictly positive growth. There is a strongly significant correlation between traffic growth and password score.

More mature sites are more likely to deploy TLS

Despite passwords being an old technology, good security engineering practices can be a process of incremental improvement. If one accepts that most sites will launch without strong password practices, then sites that have been online longer should exhibit a higher level of password security now. Using the Alexa data on sites' launch dates, we did not find a significant correlation between site age and password score, but did find that older sites are significantly more likely to deploy TLS. This gap suggests the possibility that TLS deployment may actually be seen as less important by website implementers today than it has been in the past.

Good password practices come with technical competence

We use the median load time of a site as an indicator of general technical excellence by a site's engineering team. One may expect that firms with more design skill and better business operations will implement better security practices. There is a significant positive correlation between sites loading quickly and strong password practices. A plausible explanation is that TLS deployment, guessing restrictions, and several other good password practices are much more likely to be deployed on highly customised web-servers, which are also likely to be heavily optimised to provide fast service to users.

Password security is not significantly correlated with user demographics

Alexa provides a wealth of demographic data, breaking down a site's audience by age, gender, family status, and education level. We found no significant correlations between audience demographics and password security scores. We did observe significant correlations between TLS adoption and user bases heavy on college-educated individuals and females. However, these user groups both patronise e-commerce sites more frequently, controlling for this bias removes the correlation.

5.4 Security motivations

To assess what motivates sites to implement password security measures (**RQ4**), we correlate observed password security practices with security-relevant aspects of sites' businesses. At a coarse level, we plot the distribution of password scores for our market categories, as well as sites which store payment details or offer premium accounts, in Figure 15. The significant emergent trend is that content sites are significantly behind the market average in terms of password security, while all other categories are ahead. We show further correlations for many individual password policy choices at different sites in Table 8, and find many correlations not only in overall security but also in the type of security offered by different categories of site.

Content sites trail the market significantly in security practices

Content sites score lower than the rest of the market with very strong significance. They also score significantly worse in a number of individual features, including failing to use encryption, failure to prevent guessing attacks, and being significantly more likely to send (and thus store) cleartext passwords. This trend makes sense as this segment is dominated by websites of print newspapers for which neither users nor sites have significant security requirements.

Sites storing payment details have significantly stronger security practices

Sites which store users' payment details perform significantly better in our aggregate score and in several key measures, including TLS deployment and notification to users about password reset events. These differences remain significant even when all content sites are dropped from the sample; within the generally better-performing e-commerce segment sites storing payment details perform better (many e-commerce sites choose not to store payment details). The increased motivation for password security may be in part attributable to liability if payment details are leaked (as discussed further in Section 6.3).

TLS deployment is strongly correlated with merchant facilities

Full TLS deployment, while an essential condition for password security, is only strongly adopted by e-commerce sites and particularly those storing payment details. Identity sites, including email providers and social networks, are not significantly more likely to deploy TLS. The preference for TLS at merchant sites is likely a function of the necessity of implementing TLS at least for payment processing at merchant sites, from which it is relatively easy to extend TLS to password handling. Sites that store

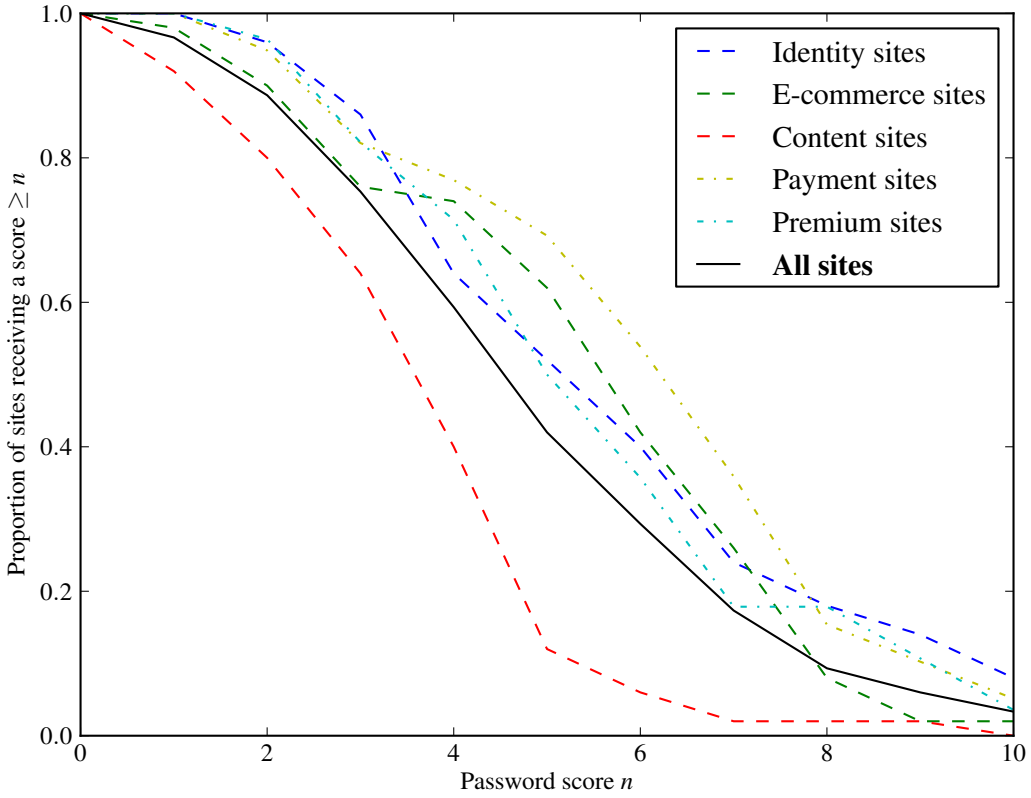


Figure 15: Cumulative distribution of password scores at different site categories and sites which store payment details or offer premium accounts. The password score is on a 10-pt scale, details are provided in § A.3. For instance, 12% percent of content sites received a password score between 5 and 10 inclusive, while only 28% of the e-commerce site scored 4 or less.

payment details are also found very strongly significant to send a notification to the registered email address once a password reset has occurred (significant for sites with merchant facilities).

Identity providers protect against weak passwords and guessing attacks

While much less likely to implement TLS, identity sites are significantly more likely to have minimum password length requirements, to prevent dictionary words, and to require the use of numbers or symbols in passwords. They are also significantly more likely to defend against guessing attacks, and significantly less likely to send cleartext passwords in email. These differences compared with e-commerce sites suggest the perceived threat model may be different, with identity providers (particularly social networks) concerned about potential guessing attacks by casual acquaintances and e-commerce sites more concerned with hacking by profit-seeking criminals.

Sites offering premium accounts do not implement significantly more security

We find no correlation suggesting that sites with premium accounts offer significantly more password security. This remains true even when sub-dividing the relatively-weak segment of content sites. Content sites which offer premium accounts (typically with access to premium content) do not perform significantly differently than content sites not offering premium content. This indicates that, perhaps due to

	Password score > median	TLS deployed correctly	Guessing attacks restricted	Minimum password length enforced	Dictionary words prohibited	Digits	Symbols	Cleartext passwords mailed	Notification of password reset	Email verified on enrolment	CAPTCHA required on enrolment
Identity segment	+	↓↓	↑		↑↑↑	+	↑	↓↓		↑	↑↑↑
E-commerce segment	↑	↑↑↑			−		−		↑	↓↓↓	↓↓↓
Content segment	↓↓↓	↓↓↓	↓		↓	−		↑	↓↓	↑↑↑	−
Premium accounts offfered					+			−			↑↑
Payment details stored	↑↑↑	↑↑↑	+	+		↑			↑↑↑	↓↓↓	−
E-mail provided	+					+	↑↑	−		−	↑↑↑
Social networking features		↓↓↓	↑↑	−	↑			↓		↑↑↑	↑↑

Table 8: Correlation between security requirements and observable indicators of password security, including our aggregate password score defined in § A.3. Statistical significance is indicated by (+/−) ($p \leq 0.2$), (↑/↓) ($p \leq 0.05$), (↑↑/↓↓) ($p \leq 0.01$), and (↑↑↑/↓↓↓) ($p \leq 0.001$).

the difficulty of preventing copying of web-based content, sites offering premium content do not view security of their accounts as important.

Password length requirements do not significantly vary between sites

While identity sites were more likely to require numbers or symbols or check passwords against a dictionary of known-bad passwords, we observed no significant correlations between site niches and password length requirements. This indicates that there is no agreement on what length of password is appropriate for different security arrangements, despite this being one of the few points for which published standards provide advice [26].

E-commerce sites internalise user authentication

While neither was common, OpenID and Facebook Connect were accepted at 13 sites in our study, with two sites accepting both. Of these, none were pure e-commerce sites and only one allowed users to store payment details (LiveJournal). Both trends were statistically significant. It appears sites processing payment information are unwilling to expose their system to the risk of an external identity provider.

5.5 Password deployment motivations

Eliciting the motivations behind password deployment (**RQ5**) is inherently difficult given the black-box character of our study, and there is little direct evidence provided by the sites themselves as to why they collect passwords. Password security is never explicitly marketed as a competitive advantage, with the exception of Hushmail and Swiss Mail, both of which promote themselves as secure webmail providers with claims such as being “the most secure web-based free email service in the world”." However, we do

notice several interesting trends which hint at the true motivations for password deployment, particularly at content sites which have less security incentive to deploy passwords.

Emails are ubiquitously collected amongst content sites without added value for users

Sites in the content segment collect email addresses on a mandatory basis as often as e-commerce sites. Only 1 out of 50 sites refrained from making email a mandatory input (Times of India, which offers its own webmail service). While it is not necessary to create server-managed accounts for content-only sites in the first place, it is even less obvious why email addresses are required instead of self-chosen pseudonyms (usernames) for this purpose. The predominance of email collection is even more surprising as only 2 of 50 content sites implemented socialising features, for which email addresses are used as a common baseline notification channel. Contrasting content sites with user identity sites by comparing email collection and the provision of socialising features between categories, we find a very strongly significant difference: content sites make email addresses mandatory much more often but use them to facilitate cross-user communication much less often.

Content sites secure email addresses rather than passwords

More interesting than the requirement of email addresses at content sites is that validation of email addresses through an account activation email is required significantly more often at content sites than at other email-collecting sites. 70% of content sites validate email addresses, whereas this proportion is only 58% among identity and 2% among e-commerce sites, both very strongly significant differences. This is curious because email activation is at least as complicated to implement as several other measures which content sites avoid, such as proactive password checking or sending notification of password reset. Despite the interest of content sites in securing email addresses rather than passwords, however, we received significantly less email in the months after opening accounts at content sites than at e-commerce sites. One explanation may be that content sites in fact aim to verify email addresses to justify to advertisers the size of their loyal readership rather than send marketing email.

Content sites use password enrolment as an opportunity to collect consumer profiles

Passwords are never collected as isolated data items; rather, they are a building block for a secured user profile. With more than three out of four content sites requiring personal information at account creation, the collection of marketing data such as job, income, and so forth is strongly significantly most common in this category. This is the highest proportion of all categories with strong significance. This marketing data can be used by a content site to gain a profile of its readership and to target advertisements (as newspapers may have their own relationships to local advertisers and not be fully dependent on online advertisement brokers). Along with the collection and verification of email addresses, it seems marketing data about the readership of a content site is a primary driver for password collection. For content sites, it seems that password schemes are used to motivate the collection of additional profile information at the moment of registration, rather than the use of passwords being motivated by the collection of personal data in future interaction with the site.

6 Economic interpretations

The observed market-wide failures in password security can be explained by diverging incentives of the market players. The market exhibits a twofold inefficiency which hinders optimal security allocation. First, efficiency is undermined by a tragedy of the commons where consumers' ability to memorise passwords is an overused resource. Second, and consequently, consumers re-use passwords across sites with varying security levels, at which point weak security practices at one site exercise a negative externality to other sites which have implemented higher security. We examine these two major effects and discuss possible methods to redress them, as well as alternative economic explanations for our observed data.

6.1 Password security as a tragedy of the commons

Common goods are characterised by an inability to restrict consumption either directly or indirectly through payments. Like public goods there is no exclusiveness, but unlike public goods common goods decline in value as they are consumed more intensively. Classical examples include natural resources such as parks or fishing grounds which tend to be overused and depleted in the absence of regulation.

Consumers' finite mental storage capacity for passwords is a common good from the viewpoint of website operators. Asking consumers to remember an additional password comes at no cost to a site operator, but can bring direct financial benefit from increased customer affinity and the ability to gather customer data (§ 5.5). Yet, each additional password places further demands on a user's memory, and may not bring real benefits to the user. To prevent depletion of their password memory, consumers must either reduce the burden for each individual password by choosing weaker passwords or reduce the cumulative burden by re-using passwords. The former tactic may have limited applicability as individual sites can enforce password restrictions (§ 4.3.2), but user surveys have revealed that users do consciously make this sacrifice. There is considerable empirical evidence that consumers more often take the latter approach of password re-use (§ 2.2).

6.2 Password insecurity as a negative externality

Password re-use is a consequence of excessive number of passwords requested of users as multiple websites compete for scarce memory resources. In addition to decreasing the ability of users to employ strong passwords, it makes disparate sites' security interdependent as a password leaked at one site can be used at any other site where the user has registered it, particularly as most sites will use the same email address to identify users. Attackers will rationally seek to extract passwords from the lowest-security websites and then re-use them at higher security websites. Password security is therefore a shared-defence problem [14] with the minimum-effort website affecting the actual security of all others,[14] although the problem here is even worse as many players have little stake in the resulting security level.

Thus, websites with poor password security impose a strong negative externality on sites which have implemented more security, as they dissipate a security cost without accountability in the market. Our experiments provided much data to support this hypothesis by showing the very low level of security implemented by content sites compared to sites processing payment details (§ 5.4). Specifically, content sites are only slightly less likely to implement minimal password choice restrictions or give password advice, neither of which would greatly help prevent cross-site compromise. However, they are significantly more likely to not implement TLS, store large databases of passwords in cleartext and not protect their membership list from probing attacks. These choices expose content sites as password oracles for attackers to use against user accounts at more secure sites when passwords are re-used [60].

6.3 Possible regulatory solutions

We suggest that the market for password security has failed in two ways. First, user's limited capacity to remember strong passwords is depleted in a tragedy of the commons as each website seeks to collect a password. Second, negative externalities from password re-use reduce the efficiency the markets as an allocation mechanism for security. Currently, the social optimum of password deployment is not reached. We propose regulatory fixes for each problem in turn.

Password tax

Preventing excessive requests for passwords is difficult; a global licensing approach with a legal prohibition on unnecessary password collection seems too rigid. A standard procedure to mitigate negative externalities is to put a price on generating them until their costs are fully internalised. A 'password tax' would impose a cost on websites for every password-protected account they store. This would prevent

[14] It could be argued to be a sum-of-efforts problem, but this doesn't greatly affect the analysis.

sites from using password-protected accounts when they have no or little business incentive for doing so, and would reward them instead for using a delegated protocol such as OpenID. If levying a direct tax is politically infeasible, password storage can still be effectively priced through yearly personal data statements [28], payable certificates or charged-for reminder emails.

Restricting password re-use by password segmentation

It would also be possible to impose direct restrictions on password re-use without limiting the number of websites which request passwords if low-security applications are prevented from accepting the same passwords as high-security ones. The optimum social welfare would be reached if password re-use were prevented by low-security sites accepting only weak passwords and high-security sites only strong passwords as low-security sites by definition don't require strong passwords. This could mean a 5 character-maximum at low-security sites and a 6 character minimum at high-security sites, Fewer password re-use externalities could dissipate with these restrictions.[15] This solution is an intermediate step towards barring password collection at low-security sites[16] which may be more tenable as renouncing passwords is incompatible with business tactics in the online news market(§ 5.5). However, it may be of limited practicality as preventing strong passwords imposes a significant implementation and usability cost on low-security websites and would likely further confuse users.

Liability

Correcting the externality caused by compromise at low-security sites is difficult because it is difficult to detect when this occurs.[17] There is space for further research on account compromise forensics and the ability to detect cross-site compromise attacks, presumably this could be detected using honeypot accounts [105]. If cross-site compromise were detected, liability laws could allow higher-security sites to receive compensation from a lower-security site. Over the long run this would provide further incentive for websites to avoid collecting passwords unless it was truly necessary for their business.

Technical standards

A technical solution would be wider availability of easy-to-use, security-graduated password toolkits, making it cheaper for low-security sites to implement weak passwords in a controlled manner as opposed to their own haphazardly-designed password deployments with globally-damaging behaviours such as sending passwords in cleartext. Such standard implementations could be strongly encouraged by licensed branding. Official standards should support reference implementations and give more proscriptive advice for proprietary solutions. In addition to increasing security, standardisation would be beneficial to users as it would decrease confusion caused by differences in custom implementations.

6.4 Alternative explanations

Password security and risk salience

A useful psychological framework for understanding privacy in websites is *privacy salience* [62]. This phenomena explains individuals' tendency in experiments to reveal less personal data when they are given more assurances that their privacy will be protected because they are made more aware that privacy violations are possible. It was found to explain a number of phenomena in the market for privacy in online social networks [21], where sites with better privacy were found to mention it less often. In the

[15] There still may be a negative externality as the require to remember at least two distinct passwords may cause users to choose both less securely, and indeed the two could be closely related.

[16] Taken to the extreme, low-security sites could only accept passwords from a set of size 1, which would be equivalent to an outright ban on password collection.

[17] Although, with a recent case involving Twitter, it was made public that Twitter was able to detect which sites' compromise led to the compromise of many Twitter accounts [89].

case of password security, it is possible that giving advice on how to pick a secure password may make the risk of password compromise salient and make users less willing to register with a site.

This could explain the rarity of password selection advice; if sites say nothing about how to prevent password compromise then a user is less likely to think that she is at risk of compromise and will be more willing to join the site. This may also explain the interesting trend of sites offering more advice at their password update screen than at the enrolment screen, when the salience effect is no longer problematic as users are already in a state of concern about their security. Similar to the case of social networks, where strong privacy assurances are given to privacy-aware users who seek out privacy information, websites may be more willing to help users who have voluntarily navigated to the password update page, as these users have already demonstrated that they care about the security of their password.

However, only a small number of aspects of password security are visible to non-technical users. Many, such as implementing TLS and preventing guessing attacks, are largely invisible, but sites fail to implement these measures with similar frequency to more visible measures. Thus, reducing risk salience cannot be a complete explanation for the insecurities observed in the market.

Password security as a lemons market

Because ordinary users are unlikely to spot the difference between high and low-quality password implementations, password security in websites can be modelled as a lemons market [111]. In applying this model, insecure sites can beat secure sites in the market with lower deployment costs if password security offers no advantage in gaining users. While we found considerable evidence that sites fail to promote their password security practices to users, we did find substantial variation in the market, with larger, more technically proficient sites and sites storing payment details implementing more password security. Furthermore, the most popular and fastest-growing sites tend to have better password practices. These observations suggest that the lemons-market explanation is too simplistic to explain the entire market, especially as identity and e-commerce sites may bear real costs of insecure implementations through customer complaints or payment card charge-backs. The content segment of the market may be better modelled as a lemons market, as site operators have no incentive to keep passwords secure.

7 Conclusions and perspectives

Our empirical study has provided a unique snapshot of how password security is implemented in practice which we hope will be useful for further analysis.[18] We can posit three major observations about the current state of password security:

Technical failures

Our data confirms some common assumptions that passwords are frequently deployed in an insecure manner. The large inconsistency of implementations between sites and the frequency of simple mistakes show an unanticipated level of insecurity and confusion. A widely available reference implementation would prevent developers from re-implementing similar functionality and introducing mistakes and design anomalies. This implementation could possibly be branded, so that users are aware they are going to a site which is using publicly-reviewed code to handle their passwords. Published standards for some implementation decisions (such as minimum length, reset mechanisms, etc.) may also help; current standards (§ 2.7) are out-dated and non-proscriptive. Most knowledge remains spread across years of often-conflicting academic research papers where it is not easily accessible for developers.

[18] `http://preibusch.de/publ/password-market`

Market failures

We also find strong evidence that market failure is leading to real insecurity on the web. Given the threat of cross-domain password attacks, insecure sites collecting passwords have the potential to impose a costly externality on more careful sites. Specific regulation banning some bad security practices can help, but solutions may require stronger regulation in the form of a password tax or increased liability which provide strong disincentives for sites to use password-protected accounts when they have no business reason for doing so and also encourage adoption of a delegated protocol such as OpenID.

Psychological failures

Finally, we propose that there may be large psychological barriers to change for both users and web developers who have grown accustomed to a set of practices around password collection descended from the early days of the web. Thus, a cult has arisen in the non-religious sense of the word, in that rituals have become entrenched and venerated independent of their utility on the modern web. In particular, the content sites in our study (predominantly online versions of print newspapers) have little security reason to deploy passwords at all. However, the behaviour has become normalised and it serves two important functions. First, as sharing passwords between people has been found to be a sign of trust and intimacy [103, 32], registering a password with a website may be a means establishing an intimate connection with a trusted brand. Second, registering passwords provides cover for collecting email addresses and marketing data. Users may feel more comfortable registering this data in the context of a password-protected account, though cookies could just as easily store customisation preferences.

This line of reasoning raises important psychological and behavioural economics questions to be tested experimentally. In the meantime, this perspective supports the claim [15] that deployment of an open, federated identity protocol such as OpenID will be opposed by current stakeholders on the web. Federated login not only removes sites' pretext for collecting personal data but their ability to establish a trusted relationship with users. Facebook Connect may have a greater chance of adoption, since while it also removes the trusted ritual of password enrolment, it preserves and enhances sites' ability to collect user data, and gives them the ability to market their sites by posting stories back to the user's social network. Federated identity protocols were designed to prevent users from having to trust a large number of online entities. However, if establishing a feeling of trust is the primary function of passwords in many current deployments, replacing them will be a difficult task.

Acknowledgements

The authors would like to thank Jonathan Anderson, Ross Anderson, Alastair Beresford, Richard Clayton, Saar Drimer, Markus Kuhn, and Arvind Narayanan for their support and feedback, as well as Andrew Lewis for his support and assistance with visualisations.

References

[1] John the Ripper. http://www.openwall.com/john/.

[2] Password Usage. *United States Federal Information Processing Standards Publication 112*, May 1985.

[3] Pubcookie Design Specifications. http://www.pubcookie.org/docs/specs.html, Februrary 2003.

[4] *ISO/IEC 27001:2005 - Information technology – Security techniques – Information security management systems*. International Organisation for Standardisation, October 2005.

[5] OpenID Authentication 2.0 - Draft 11. http://openid.net/specs/openid-authentication-2_0-11.html, January 2007.

[6] Alexa: The Web Information Company, Feb 2010.

[7] Diceware Passphrase Generator. http://www.diceware.com, 2010.

[8] Facebook Connect. *http://www.facebook.com/advertising/?connect*, 2010.

[9] KeePass. http://en.wikipedia.org/wiki/KeePass, 2010.

[10] Microsoft Passport, Feb 2010.

[11] PasswordSafe. http://www.passwordsafe.com/, Feb 2010.

[12] Verified by Visa. www.visa.com/verifiedbyvisa/, 2010.

[13] Anne Adams and Martina Angela Sasse. Users are Not the Enemy. *Commun. ACM*, 42(12):40–46, 1999.

[14] Ross Anderson and Tyler Moore. Information Security Economics – and Beyond . *Advances in Cryptology - CRYPTO 2007*, 2007.

[15] Michael Arrington. Is OpenID Being Exploited By The Big Internet Companies? *TechCrunch*, March 2008.

[16] Farzganeh Asgharpour and Markus Jakobsson. Adaptive Challenge Questions Algorithm in Password Reset/Recovery. Presented at First International Workshop on Security for Spontaneous Interaction (IWIISI), 2007.

[17] Walter Belgers. UNIX Password Security, 1993.

[18] Steven M. Bellovin and Michael Merritt. Encrypted Key Exchange: Password-Based Protocols SecureAgainst Dictionary Attacks. In *SP '92: Proceedings of the 1992 IEEE Symposium on Security and Privacy*, page 72, Washington, DC, USA, 1992. IEEE Computer Society.

[19] Steven M. Bellovin and Michael Merritt. Augmented encrypted key exchange: a password-based protocol secure against dictionary attacks and password file compromise. In *CCS '93: Proceedings of the 1st ACM conference on Computer and communications security*, pages 244–250, New York, NY, USA, 1993. ACM.

[20] Joseph Bonneau, Mike Just, and Greg Matthews. What's in a Name: Evaluating Statistical Attacks Against Personal Knowledge Questions. *Financial Crypto '10: The Fourteenth International Conference on Financial Cryptography and Data Security*, 2010.

[21] Joseph Bonneau and Sören Preibusch. The Privacy Jungle: On the Market for Data Protection in Social Networks. In *The Eighth Workshop on the Economics of Information Security (WEIS 2009)*, 2009.

[22] Andrew Bortz and Dan Boneh. Exposing private information by timing web applications. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 621–628, New York, NY, USA, 2007. ACM.

[23] Sacha Brostoff and Angela Sasse. Ten strikes and you're out: Increasing the number of login attempts can improve password usability. In *Proceedings of CHI 2003 Workshop on HCI and Security Systems*. John Wiley, 2003.

[24] Bundesamt für Sicherheit in der Informationstechnik (BSI). *BSI-Standard 100-2. IT-Grundschutz Methodology*. 2 edition, 2008.

[25] Bundesamt für Sicherheit in der Informationstechnik (BSI) (Federal Office for Information Security). *IT-Grundschutz Catalogues*. 2005.

[26] William E. Burr, Donna F. Dodson, and W. Timothy Polk. Electronic Authentication Guideline. *NIST Special Publication 800-63*, April 2006.

[27] Joseph A. Cazier and B. Dawn Medlin. Password Security: An Empirical Investigation into E-Commerce Passwords and Their Crack Times. *Information Systems Security*, 15(6):45–55, 2006.

[28] Chaos Computer Club (CCC). Datenbrief. `http://www.ccc.de/datenbrief`, January 2010.

[29] Richard M. Conlan and Peter Tarasewich. Improving interface designs to help users choose better passwords. In *CHI '06: CHI '06 extended abstracts on Human factors in computing systems*, pages 652–657, New York, NY, USA, 2006. ACM.

[30] Baris Coskun and Cormac Herley. Can "Something You Know" Be Saved? In *ISC '08: Proceedings of the 11th international conference on Information Security*, pages 421–440, Berlin, Heidelberg, 2008. Springer-Verlag.

[31] Johanna Bromberg Craig, Wes Craig, Kevin McGowan, and Jarod Malestein. The Cosign Web Single Sign-On Scheme. `http://cosign.sourceforge.net/media/cosignscheme2006a.rtf`, June 2006.

[32] danah boyd. answers to questions from Twitter on teen practices. *apophenia*, April 2009.

[33] Darren Davis, Fabian Monrose, and Michael K. Reiter. On User Choice in Graphical Password Schemes. In *SSYM'04: Proceedings of the 13th conference on USENIX Security Symposium*, pages 11–11, Berkeley, CA, USA, 2004. USENIX Association.

[34] Paul Dourish, E. Grinter, Jessica Delgado de la Flor, and Melissa Joseph. Security in the Wild: User Strategies for Managing Security as an Everyday, Practical Problem. *Personal Ubiquitous Comput.*, 8(6):391–401, 2004.

[35] Laura Falk, Atul Prakash, and Kevin Borders. Analyzing websites for user-visible security design flaws. In *SOUPS '08: Proceedings of the 4th symposium on Usable privacy and security*, pages 117–126, New York, NY, USA, 2008. ACM.

[36] David C. Feldmeier and Philip R. Karn. UNIX Password Security - Ten Years Later. In *CRYPTO '89: Proceedings of the 9th Annual International Cryptology Conference on Advances in Cryptology*, pages 44–63, London, UK, 1990. Springer-Verlag.

[37] Dinei Florêncio and Cormac Herley. A large-scale study of web password habits. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 657–666, New York, NY, USA, 2007. ACM.

[38] Dinei Florêncio and Cormac Herley. One-Time Password Access to Any Server without Changing the Server. In *ISC '08: Proceedings of the 11th international conference on Information Security*, pages 401–420, Berlin, Heidelberg, 2008. Springer-Verlag.

[39] Dinei Florêncio, Cormac Herley, and Baris Coskun. Do strong web passwords accomplish anything? In *HOTSEC'07: Proceedings of the 2nd USENIX workshop on Hot topics in security*, pages 1–6, Berkeley, CA, USA, 2007. USENIX Association.

[40] John Franks, Phillip M. Hallam-Baker, Jeffery L. Hostetler, Scott D. Lawrence, Paul J. Leach, Ari Luotonen, and Lawrence C. Stewart. RFC2617: HTTP Authentication: Basic and Digest Access Authentication. 1999.

[41] Kevin Fu, Emil Sit, Kendra Smith, and Nick Feamster. Dos and DonâĂŹts of Client Authentication on the Web. Technical Report 818, MIT, 2001.

[42] Steven Furnell. An assessment of website password practices. *Computers & Security*, 26(7-8):445–451, 2007.

[43] Eran Gabber, Phillip B. Gibbons, Yossi Matias, and Alain J. Mayer. How to Make Personalized Web Browising Simple, Secure, and Anonymous. In *FC '97: Proceedings of the First International Conference on Financial Cryptography*, pages 17–32, London, UK, 1997. Springer-Verlag.

[44] Simson L. Garfinkel. Email-Based Identification and Authentication: An Alternative to PKI? *IEEE Security and Privacy*, 1(6):20–26, 2003.

[45] Shirley Gaw and Edward W. Felten. Password Management Strategies for Online Accounts. In *SOUPS '06: Proceedings of the Second Symposium on Usable Privacy and Security*, pages 44–55, New York, NY, USA, 2006. ACM.

[46] Oded Goldreich and Yehuda Lindell. Session-Key Generation Using Human Passwords Only. *J. Cryptol.*, 19(3):241–340, 2006.

[47] Philippe Golle and David Wagner. Cryptanalysis of a Cognitive Authentication Scheme (Extended Abstract). In *SP '07: Proceedings of the 2007 IEEE Symposium on Security and Privacy*, pages 66–70, Washington, DC, USA, 2007. IEEE Computer Society.

[48] Joshua Gomez, Travis Pinnick, and Ashkan Soltani. KnowPrivacy. Technical Report 2009-037, University of California at Berkeley, 2009.

[49] Virgil Griffith and Markus Jakobsson. Messin' with Texas: Deriving Mother's Maiden Names Using Public Records. *Applied Cryptography and Network Security*, 2005.

[50] William J. Haga and Moshe Zviran. Question-and-answer passwords: an empirical evaluation. *Inf. Syst.*, 16(3):335–343, 1991.

[51] J. Alex Halderman, Brent Waters, and Edward W. Felten. A convenient method for securely managing passwords. In *WWW '05: Proceedings of the 14th international conference on World Wide Web*, pages 471–479, New York, NY, USA, 2005. ACM.

[52] Neil Haller. The S/KEY One-Time Password System. *Proceedings of the ISOC Symposium on Network and Distributed System Security*, 1994.

[53] Neil Haller, Craig Metz, Philip J. Nesser II, and Mike Straw. *RFC 2289: A One-Time Password System*. IETF, February 1998.

[54] Eran Hammer-Lahav and David Recordon. The OAuth 1.0 Protocol. `http://tools.ietf.org/html/draft-hammer-oauth-10`, February 2010.

[55] Cormac Herley and Dinei Florêncio. Protecting Financial Institutions from Brute-Force Attacks. In *Proceedings of The Ifip Tc 11 23rd International Information Security Conference*, pages 681–685, New York, NY, USA, 2008. Springer.

[56] Cormac Herley, Paul C. Oorschot, and Andrew S. Patrick. Passwords: If We're So Smart, Why Are We Still Using Them? pages 230–237, 2009.

[57] Laurie J. Heyer, Semyon Kruglyak, and Shibu Yooseph. Exploring expression data: identification and analysis of coexpressed genes. *Genome research*, 9(1), 1999.

[58] Nicholas J. Hopper and Manuel Blum. Secure Human Identification Protocols. In *ASIACRYPT '01: Proceedings of the 7th International Conference on the Theory and Application of Cryptology and Information Security*, pages 52–66, London, UK, 2001. Springer-Verlag.

[59] Trusteer Inc. Reused Login Credentials. February 2010.

[60] Blake Ives, Kenneth R. Walsh, and Helmut Schneider. The Domino Effect of Password Reuse. *Commun. ACM*, 47(4):75–78, 2004.

[61] Ian Jermyn, Alain Mayer, Fabian Monrose, Michael K. Reiter, and Aviel D. Rubin. The design and analysis of graphical passwords. In *SSYM'99: Proceedings of the 8th conference on USENIX Security Symposium*, pages 1–1, Berkeley, CA, USA, 1999. USENIX Association.

[62] Leslie K. John, Alessandro Acquisti, and George Loewenstein. The Best of Strangers: Context Dependent Willingness to Divulge Personal Information . *Available at SSRN: http://ssrn.com/abstract=1430482*, 2009.

[63] Mike Just and David Aspinall. Personal Choice and Challenge Questions: A Security and Usability Assessment. In *SOUPS '09: Proceedings of the Fifth Symposium on Usable Privacy and Security*, 2009.

[64] Chris Karlof, J. D. Tygar, and David Wagner. A user study design for comparing the security of registration protocols. In *UPSEC'08: Proceedings of the 1st Conference on Usability, Psychology, and Security*, pages 1–14, Berkeley, CA, USA, 2008. USENIX Association.

[65] Jonathan Katz, Rafail Ostrovsky, and Moti Yung. Efficient Password-Authenticated Key Exchange Using Human-Memorable Passwords. In *EUROCRYPT '01: Proceedings of the International Conference on the Theory and Application of Cryptographic Techniques*, pages 475–494, London, UK, 2001. Springer-Verlag.

[66] Daniel Klein. "Foiling the Cracker":A Survey of, and Improvements to, Password Security. In *Proceedings of the 2nd USENIX Security Workshop*, pages 5–14, 1990.

[67] John Kohl and B. Clifford Neuman. *RFC 1510: The Kerberos Network Authentication Service (V5)*. IETF, November 1993.

[68] David P. Kormann and Aviel D. Rubin. Risks of the Passport single signon protocol. *Computer Networks*, 33(1-6):51–58, 2000.

[69] Balachander Krishnamurthy and Craig E. Wills. On the leakage of personally identifiable information via online social networks. In *WOSN '09: Proceedings of the 2nd ACM workshop on Online social networks*, pages 7–12, New York, NY, USA, 2009. ACM.

[70] Markus Kuhn. OTPW — a one-time password login package. `http://www.cl.cam.ac.uk/~mgk25/otpw.html`, 1998.

[71] Cynthia Kuo, Sasha Romanosky, and Lorrie Faith Cranor. Human Selection of Mnemonic Phrase-based Passwords. In *SOUPS '06: Proceedings of the second symposium on Usable privacy and security*, pages 67–78, New York, NY, USA, 2006. ACM.

[72] Leslie Lamport. Password authentication with insecure communication. *Commun. ACM*, 24(11):770–772, 1981.

[73] Ben Laurie. OpenID: Phishing Heaven. `http://www.links.org/?p=187`, January 2007.

[74] Nancy J. Lightner. What users want in e-commerce design: effects of age, education and income. *Ergonomics*, 46(1):153–168, 2003.

[75] Mohammad Mannan and Paul C. van Oorschot. Security and usability: the gap in real-world online banking. In *NSPW '07: Proceedings of the 2007 Workshop on New Security Paradigms*, pages 1–14, New York, NY, USA, 2008. ACM.

[76] Drew Mazurek. Central Authentication Service Protocol. `http://www.jasig.org/cas/protocol`, May 2005.

[77] John H. McDonald. *Handbook of Biological Statistics*. Sparky House Publishing, Baltimore, Maryland, 2nd edition, 2009.

[78] Ari Medvinsky and Matthew Hur. *RFC 2712: Addition of Kerberos Cipher Suites to Transport Layer Security (TLS)*. IETF, October 1999.

[79] R. L. "Bob" Morgan, Scott Cantor, Steven Carmody, Walter Hoehn, and Ken Klingenstein. Federated Security: The Shibboleth Approach. *EDUCAUSE quarterly*, 27(4), 2004.

[80] Robert Morris and Ken Thompson. Password security: a case history. *Commun. ACM*, 22(11):594–597, 1979.

[81] Steven Murdoch and Ross Anderson. Verified by Visa and MasterCard SecureCode: or, How Not to Design Authentication. *Financial Crypto '10: The Fourteenth International Conference on Financial Cryptography and Data Security*, 2010.

[82] Arvind Narayanan and Vitaly Shmatikov. Fast dictionary attacks on passwords using time-space tradeoff. In *CCS '05: Proceedings of the 12th ACM conference on Computer and communications security*, pages 364–372, New York, NY, USA, 2005. ACM.

[83] Donald A. Norman. When Security Gets in the Way. *Interactions*, 16(6):60–63, 2009.

[84] Gilbert Notoatmodjo and Clark Thomborson. Passwords and Perceptions. In Ljiljana Brankovic and Willy Susilo, editors, *Seventh Australasian Information Security Conference (AISC 2009)*, volume 98 of *CRPIT*, pages 71–78, Wellington, New Zealand, 2009. ACS.

[85] Philippe Oechslin. Making a Faster Cryptanalytic Time-Memory Trade-Off . *Advances in Cryptology - CRYPTO 2003*, 2003.

[86] Rolf Oppliger. Microsoft .NET Passport: A Security Analysis. *Computer*, 36(7):29–35, 2003.

[87] Andreas Pashalidis and Chris J. Mitchell. *A Taxonomy of Single Sign-On Systems* , volume 2727 of *Information Security and Privacy*, pages 219–235. Springer, 2003.

[88] Rachael Pond, John Podd, Julie Bunnell, and Ron Henderson. Word Association Computer Passwords: The Effect of Formulation Techniques on Recall and Guessing Rates. *Computers & Security*, 19(7):645–656, 2000.

[89] Brian Prince. Twitter Details Phishing Attacks Behind Password Reset. *eWeek*, January 2010.

[90] Ariel Rabkin. Personal knowledge questions for fallback authentication: security questions in the era of Facebook. In *SOUPS '08: Proceedings of the 4th symposium on Usable privacy and security*, pages 13–23, New York, NY, USA, 2008. ACM.

[91] David Recordon and Dick Hardt. The OAuth 2.0 Protocol. `http://tools.ietf.org/html/draft-hammer-oauth2-00`, April 2010.

[92] David Recordon and Drummond Reed. OpenID 2.0: a platform for user-centric identity management. In *DIM '06: Proceedings of the second ACM workshop on Digital identity management*, pages 11–16, New York, NY, USA, 2006. ACM.

[93] Shannon Riley. Password Security: What Users Know and What They Actually Do. *Usability News*, 8(1), 2006.

[94] Blake Ross, Collin Jackson, Nick Miyake, Dan Boneh, and John C. Mitchell. Stronger password authentication using browser extensions. In *SSYM'05: Proceedings of the 14th conference on USENIX Security Symposium*, pages 2–2, Berkeley, CA, USA, 2005. USENIX Association.

[95] Aviel D. Rubin. Independent one-time passwords. In *SSYM'95: Proceedings of the 5th conference on USENIX UNIX Security Symposium*, pages 15–15, Berkeley, CA, USA, 1995. USENIX Association.

[96] Martina Angela Sasse, Sacha Brostoff, and Dirk Weirich. Transforming the 'Weakest Link' — a Human/Computer Interaction Approach to Usable and Effective Security. *BT Technology Journal*, 19(3):122–131, 2001.

[97] Stuart Schechter, A. J. Bernheim Brush, and Serge Egelman. It's No Secret. Measuring the Security and Reliability of Authentication via "Secret” Questions. In *SP '09: Proceedings of the 2009 30th IEEE Symposium on Security and Privacy*, pages 375–390, Washington, DC, USA, 2009. IEEE Computer Society.

[98] Stuart Schechter, Serge Egelman, and Robert W. Reeder. It's not what you know, but who you know: a social approach to last-resort authentication. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 1983–1992, New York, NY, USA, 2009. ACM.

[99] Stuart E. Schechter, Rachna Dhamija, Andy Ozment, and Ian Fischer. EmperorâĂŹs new security indicators: An evaluation of website authentication and the effect of role playing on usability studies. In *In Proceedings of the 2007 IEEE Symposium on Security and Privacy*, 2007.

[100] Roland Schemers and Russ Allbery. WebAuth V3 Technical Specification. `http://webauth.stanford.edu/protocol.html`, 2009.

[101] Bruce Schneier. Real-World Passwords, December 2006.

[102] Richard Shay, Saranga Komanduri, Patrick Gage Kelley, Pedro Giovanni Leon, Michelle L. Mazurek, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. Encountering Stronger Password Requirements: User Attitudes and Behaviors. In *SOUPS '10: Proceedings of the Sixth Symposium on Usable privacy and Security*. ACM, 2010.

[103] Supriya Singh, Anuja Cabraal, Catherine Demosthenous, Gunela Astbrink, and Michele Furlong. Password Sharing: Implications for Security Design Based on Social Practice. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 895–904, New York, NY, USA, 2007. ACM.

[104] Eugene Spafford. Observations on Reusable Password Choices. In *Proceedings of the 3rd USENIX Security Workshop*, 1992.

[105] Lance Spitzner. *Honeypots: Tracking Hackers*. Addison-Wesley, 2003.

[106] Jennifer G. Steiner, Clifford Neuman, and Jeffrey I. Schiller. Kerberos: An Authentication Service for Open Network Systems. *USENIX Winter Conference*, pages 191–202, 1988.

[107] David Taylor, Thomas Wu, and Trevor Perrin. *RFC 5054: Using the Secure Remote Password (SRP) Protocol for TLS Authentication*. IETF, November 2007.

[108] Julie Thorpe and Paul C. van Oorschot. Human-Seeded Attacks and Exploiting Hot-Spots in Graphical Passwords. In *SS'07: Proceedings of 16th USENIX Security Symposium*, Berkeley, CA, USA, 2007. USENIX Association.

[109] Tim Valentine. An Evaluation of the Passfaces Personal Authentication System . Technical report, Goldmsiths College University of London, 1998.

[110] Ashlee Vance. If Your Password Is 123456, Just Make It HackMe . *The New York Times*, January 2010.

[111] Tony Vila, Rachel Greenstadt, and David Molnar. Why We Can't Be Bothered to Read Privacy Policies: Models of Privacy Economics as a Lemons Market. In *ICEC '03: Proceedings of the 5th International Conference on Electronic commerce*, pages 403–407, New York, NY, USA, 2003. ACM.

[112] Jon Warbrick. The Cambridge Web Authentication System: WAA->WLS communication protocol. `http://raven.cam.ac.uk/project/waa2wls-protocol.txt`, November 2005.

[113] Daphna Weinshall. Cognitive Authentication Schemes Safe Against Spyware (Short Paper). In *SP '06: Proceedings of the 2006 IEEE Symposium on Security and Privacy*, pages 295–300, Washington, DC, USA, 2006. IEEE Computer Society.

[114] Matt Weir, Sudhir Aggarwal, Breno de Medeiros, and Bill Glodek. Password Cracking Using Probabilistic Context-Free Grammars. In *SP '09: Proceedings of the 2009 30th IEEE Symposium on Security and Privacy*, pages 391–405, Washington, DC, USA, 2009. IEEE Computer Society.

[115] Dirk Weirich and Martina Angela Sasse. Pretty good persuasion: a first step towards effective password security in the real world. In *NSPW '01: Proceedings of the 2001 workshop on New security paradigms*, pages 137–143, New York, NY, USA, 2001. ACM.

[116] Susan Wiedenbeck, Jim Waters, Jean-Camille Birget, Alex Brodskiy, and Nasir Memon. PassPoints: design and longitudinal evaluation of a graphical password system. *Int. J. Hum.-Comput. Stud.*, 63(1-2):102–127, 2005.

[117] Thomas Wu. The secure remote password protocol. *Proceedings of the 1998 Internet Society Network and Distributed System Security Symposium*, 1998.

[118] Jeff Yan, Alan Blackwell, Ross Anderson, and Alasdair Grant. Password Memorability and Security: Empirical Results. *IEEE Security and Privacy Magazine*, 2(5):25, 2004.

A.1 Complete list of sites

Our sample consisted of 150 sites, with 50 coming from each of three categories. The complete list is shown in alphabetical order within each category in Table 9.

A.2 Technical details of evaluation setup

Because websites may tailor interaction details based on any observable data about the user, we were careful to keep the evaluation conditions constant. All enrolment was done using Mozilla Firefox v 3.5.8 running on Ubuntu 9.10 Linux, configured to accept all cookies. All traffic was sent from IP addresses in the Cambridge Computer Laboratory's address space `128.232.*.*`. Data collection was carried out in February 2010. All websites received the same form details whenever possible. A standard (bogus) identity was used including a non-existent address in California, a non-existent telephone number, and a birthdate of Dec. 12, 1972. All mailing list offers were opted-in to.

We initialised our browser using a brand new profile which was used for no purpose except signing up for the sites in this study. We made use of the following browser extensions to automate signup and data collection (though maintained a human in the loop at all times):

- **Autofill Forms 0.9.5.2** for automatically filling in forms with consistent user data.
- **CipherFox 2.3.0** for examining server certificate information.
- **Cookie Monster 0.98.0** for examining and modifying cookies.
- **DOM Inspector 2.0.4** for examining the structure of websites.
- **Greasemonkey 0.8.20100211.5** for running automated scripts.
- **Screengrab 0.96.2** for capturing screen shots of surveyed sites.
- **Tamper Data 11.0.1** for examining form data submitted.

A.3 Password security score

The password security score is intended to give a rough estimate of the overall level of password security implemented by a website. The numerical formula is given in Table 10. The score only reflects the number of steps sites have made to attempt to increase security for analysis purposes, it is not intended as absolute rating of good security. Graphical distributions of scores are shown in Figures 14 and 15.

A.4 Condensed password policy tuples

The password policy score is intended to capture a site's password security policy in a concise tuple of 10 categorical features. The numerical formula is given in Table 11. Each feature captures one general area of password security policy. The values are different for each feature and are categorical in all cases except for the minimum length of allowable passwords. Unlike the password security score, the policy tuples cannot be ranked, they can only be compared to each other by Hamming distance.

A.5 Clustering details

Using our password policy tuples (as defined in § A.4), we clustered the sites using the Quality Threshold (QT) clustering algorithm [57]. The results are shown in Figure 13. Our distance metric was Hamming distance: any disagreed policy elements increasing the distance by 1 between a pair of sites. Thus, any two policies are an integral distance apart in $[0, 10]$. Our measure of cluster quality was the radius of the cluster, defined as the maximum distance from the centroid element of the cluster to any other point. We obtained the best results using a quality threshold (maximal cluster radius) of 3.

Identity	Content	E-commerce
3Dup	Ask Jeeves	Abercrombie & Fitch
aNobii	Bill O'Reilly Online	AliBaba
AOL	Bloomberg	Amazon
Associated Content	Canada.com	Anthropologie
Bodybuilding.com	Chicago Tribune	Art Beads
CBS Sports	CNBC	Barnes & Noble
Craigslist	CNET	Bath & Body Works
Deviant Art	CNN	Best Buy
Digg	Fairfax Digital	Blick
Ebay	Financial Times	Build-A-Bear Workshop
EmailAccount	Forbes	Buy.com
ESPN	Fort Worth Star-Telegram	Cafe Press
Facebook	Houston Chronicle	CD Wow
Fertility Friend	Indian Express	Costco
Gamespot	LA Times	DVD Empire
Gawab	Miami.com	eBooks
Godmail	Milwaukee Journal Sentinel	Eddie Bauer
Google	Nashville Scene	efollet.com
hi5	New York Post	Football Fanatics
Huffington Post	New York Times	Frederick's of Hollywood
Hushmail	On The Snow	Gap
IMDB	Orlando Sentinel	Hermes
Last.fm	philly.com	Home Depot
LinkedIn	PhillyBurbs	Horchow
LiveJournal	Reuters	IKEA
LiveMocha	Sacramentro Bee	JCPenney
Mail.com	San Francisco Chronicle	Lucky Vitamin
Mail2World	San Jose Mercury News	Macy's
Microsoft Live	Seattle Weekly	NewEgg
Mixx	Sustainable Business	Oriental Trading
MySpace	TCPalm	Overstock
PhotoBucket	The Courier-Journal	Rand McNally
Plaxo	The Dallas Morning News	REI
Reddit	The Drum	Sears
rediff	The Economist	Sephora
ResearchGate	The Florida-Times Union	ShopBop
SoftHome	The Guardian	Shoplet
Sonico	The Lincoln Journal Star	Sierra Trading Post
StumbleUpon	The Post-Tribune	Spiegel
Swiss Mail	The Press-Telegram	Target
TalkBizNow	The Tennessean	The Children's Place
The Pirate Bay	The Topeka Capital-Journal	The Golf World
Topix	The Weather Channel	Things Remembered
Twitter	Times of India	Ticket Web
Wasabi	Times Online	TicketMaster
Wikipedia	Truthdig	TigerDirect
Wordpress	USA Today	Two Peas in a Bucket
Xanga	Wall Street Journal	Victoria's Secret
Yahoo!	Washington Post	Walmart
ZapZone Network	Weather Underground	Zappos!

Table 9: Complete list of sites surveyed

section	feature	scoring
	enrolment	
4.3.1	Password selection advice given	$+1$ pt
4.3.2	Minimum password length required	$+1$ pt
4.3.2	Dictionary words prohibited	$+1$ pt
4.3.2	Numbers or symbols required	$+1$ pt
4.9	User list protected from probing	$+1$ pt
4.2.2	Cleartext password sent in email after enrolment	-1 pt
	login	
4.4.2	Password hashed in-browser before POST	$+1$ pt
4.8	Limits placed on password guessing	$+1$ pt
4.9	User list protected from probing	$+1$ pt
4.5	Federated identity login accepted	$+1$ pt
	password update	
4.6	Password re-entry required to authorise update	$+1$ pt
4.6	Notification email sent after password reset	$+1$ pt
	password recovery	
4.7	Password update required after recovery	$+1$ pt
4.7.1	Cleartext password sent in email upon request	-1 pt
4.9	User list protected from probing	$+1$ pt
	encryption	
4.10	Full TLS for all password submission	$+2$ pts
4.10	POST only TLS for password submission	$+1$ pt

Table 10: Password score criteria, annotated with the section numbers describing each criteria in more detail. Scores could possibly fall in the range $[-2, 13]$, we actually observed scores only in the range $[0, 10]$, so we consider the scoring to be on a ten-point scale. Websites receiving a full 10 points included eBay, IKEA, LiveJournal, Microsoft Live, and Yahoo!. Graphical distributions of scores are shown in Figures 14 and 15.

§	feature	possible values	cardinality
4.2	Enrolment email contents	Email verification$_{\{Y,N\}}$ $\times$ Username$_{\{Y,N\}}$ $\times$ Cleartext password$_{\{Y,N\}}$	8
4.3.1	Password advice	Hard to guess$_{\{Y,N\}}$ $\times$ Not a common word$_{\{Y,N\}}$ $\times$ Use special characters$_{\{Y,N\}}$ $\times$ Graphical strength indicator$_{\{Y,N\}}$	16
4.3.2	Minimum password length	$\mathbb{Z}_8$	8
4.3.2	Password requirements	Non-dictionary$_{\{Y,N\}}$ $\times$ Contains numbers$_{\{Y,N\}}$ $\times$ Contains symbols$_{\{Y,N\}}$ $\times$ Not previously used $_{\{Y,N\}}$	16
4.5	Federated login support	OpenID provided$_{\{Y,N\}}$ $\times$ OpenID accepted$_{\{Y,N\}}$ $\times$ Facebook Connect accepted$_{\{Y,N\}}$	16
4.6	Password update	Password re-entry required$_{\{Y,N\}}$ $\times$ Notification sent$_{\{Y,N\}}$ $\times$ Required after recovery$_{\{Y,N\}}$	8
4.7	Password recovery mechanism	Personal Knowledge$_{\{Y,N\}}$ $\times$ {On-screen, cleartext email, temporary password, reset link}	8
4.8	Brute force restrictions	{$\varnothing$, CAPTHCA, timeout, forced reset}	4
4.9	User probing restricted	Enrolment$_{\{CAPTCHA,N\}}$ $\times$ Login$_{\{Y,N\}}$ $\times$ Reset$_{\{Y,CAPTCHA,N\}}$	12
4.10	TLS deployment	{Full, POST-only, mixed,$\varnothing$}	4

Table 11: Features in condensed password policy tuples, annotated with the section numbers describing each feature in more detail. The set of possible values for each feature is noted. Technically, the minimum password length can be any value in $\mathbb{Z}$, which would give a cardinality of ∞, but we only observed values in $[1, 8]$.