Email-based password recovery — risking or rescuing users?

Fatma Al Maqbali and Chris J Mitchell
Information Security Group, Royal Holloway, University of London
fatmaa.soh@cas.edu.om, me@chrismitchell.net

***Abstract*—Secret passwords are very widely used for user authentication to websites, despite their known shortcomings. Most websites using passwords also implement password recovery to allow users to re-establish a shared secret if the existing value is forgotten; many such systems involve sending a password recovery email to the user, e.g. containing a secret link. The security of password recovery, and hence the entire user-website relationship, depends on the email being acted upon correctly; unfortunately, as we show, such emails are not always designed to maximise security and can introduce vulnerabilities into recovery. To understand better this serious practical security problem, we surveyed password recovery emails for 50 of the top English language websites. We investigated a range of security and usability issues for such emails, covering their design, structure and content (including the nature of the user instructions), the techniques used to recover the password, and variations in email content from one web service to another. Many well-known web services, including Facebook, Dropbox, and Microsoft, suffer from recovery email design, structure and content issues. This is, to our knowledge, the first study of its type reported in the literature. This study has enabled us to formulate a set of recommendations for the design of such emails.**

***Index Terms*—password recovery, email-based password recovery, content and design of email-based password recovery.**

I. Introduction

Password-based authentication is widely used; 90% of the 1000 most-visited websites authenticate users with a textual password [1]. Users will often have dozens of username/password pairs to remember, a very challenging requirement. As a result, users commonly forget or mislay passwords. Thus websites typically offer a password recovery system to allow legitimate users to recover account access by redistributing the old password or creating a new one.

Websites often offer multiple password recovery options, e.g. based on security questions, a registered email address, or a mobile number. Email-based password recovery is widely implemented [2], [3], and is likely to remain in wide use for years to come. In a typical such system, when a user initiates password recovery a *password recovery email* is sent to the user's registered email address. The recovery email, see e.g. Figure 1, will typically contain instructions on how to recover the password; these instructions might include clicking on a link or entering a verification code or temporary password.

Since email-based password recovery is widely used, it is important to understand whether it is secure in practice. Unfortunately, as we show, some recovery emails are not well-designed and can introduce vulnerabilities into the recovery process. In this paper, we analyse password recovery emails by examining their content and design, and use this analysis to make recommendations regarding their design. To our knowledge there are no published studies on the design of recovery emails, despite the potential of poor design to give rise to serious vulnerabilities. This observation has motivated the work described here.

To understand better the size and nature of this serious practical security problem, we manually surveyed password recovery emails for 50 of the top English language websites[1]. The study examined the content of recovery emails and evaluated their security and usability, including: their ease of use, their overall design, the clarity of instructions, and the techniques used to recover the password.

The paper is structured as follows. Section II is a review of previous work on password recovery and email-based password recovery. In section III we review the use of password recovery emails; Section IV then reviews the risks associated with email-based password recovery, which forms the basis for our subsequent analysis. Section V gives the scope and methodology for the study, and Section VI provides the main findings from our survey. In Section VII we use these findings to provide recommendations on the design of email-based password recovery. Section VIII concludes the paper, including a discussion of possible directions for further research.

II. Previous work

Bonneau et al. [2] examined 150 web sites which offer free services and found that 92% of them offer email-based password recovery. This means that ensuring the recovery emails are well-designed is of great practical importance for security. A number of authors have examined issues relating to email-based web password recovery. In 2003, Garfinkel [3] discussed email-based identification and authentication and described ways of improving its security. More recently, Al Maqbali and Mitchell [4] discussed the security issues arising from the password recovery process, including from the use of emails, although they did not focus on the design of the emails themselves. Specific aspects of password recovery systems, notably the design and use of Short Message Service (SMS)-based password recovery, were examined by Gelernter et al. [5], who observed serious vulnerabilities arising from poor

[1] The list was taken from https://majestic.com/reports/majestic-million.

978-1-5386-7931-9/18/$31.00 © 2018 IEEE

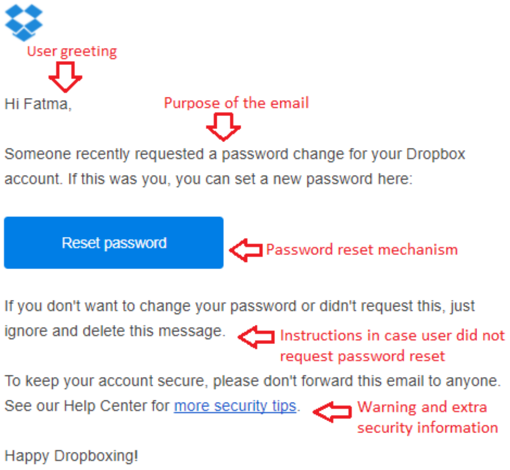
design of such SMSs. Just as in SMS-based recovery, the use of emails for password recovery has inherent problems, [4], [6]. E.g., the email may fall into the wrong hands if the email address used is no longer correct. However, our main concern here is how the design of recovery emails can affect security.

III. Password Recovery Emails

A. Use of recovery emails

Email-based password recovery (see, for example, [4]) involves an email being sent to the user by the web service; this email contains a secret which, on the assumption that the email can only reach the intended user, is used to authorise the password recovery request. The email may also contain information which helps the user to reset their password at the website. These emails vary in content and design, although in all cases it is vital that the email is not available to third parties who could use the secret in the email to hijack the recovery process. Given that correct use of the email by its recipient is security-critical, the content and design of the email are clearly of importance in practice, since they will affect how the email is treated by the user.

Figure 1 shows a Dropbox password recovery email. Such emails typically include a user greeting, the email purpose, instructions (possibly including what to do if the user did not request recovery), and contact details. The instructions will include a (one-time-secret) mechanism enabling password recovery, e.g. a URL (link) or code/temporary password. Clicking on such a link redirects the browser to a page enabling the user to set up a new password. Verification codes/temporary passwords typically give temporary access to a user account purely to establish a new password, e.g. as used by Amazon and Wikipedia.



User greeting

Hi Fatma,

Purpose of the email

Someone recently requested a password change for your Dropbox account. If this was you, you can set a new password here:

Reset password

Password reset mechanism

If you don't want to change your password or didn't request this, just ignore and delete this message.

Instructions in case user did not request password reset

To keep your account secure, please don't forward this email to anyone. See our Help Center for more security tips.

Warning and extra security information

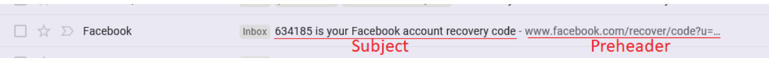
Happy Dropboxing!

Fig. 1. Example password recovery email (from dropbox.com)

B. Email structure

In general, emails can be divided into *header* and *body* [7]; however, for our purposes we further divide password recovery emails into four components: the email preheader, email header, email body and email signature (where preheader and signature are parts of the body), as follows.

- **Email preheader (Johnson Box[2] or Snippet)**. This corresponds to the first few words of an email, the exact length being determined by the email client. Figure 2 shows examples of Gmail preheaders. It is typically displayed by the email client after the subject header field (see below), to give the recipient an idea of what the email is about[3]. Despite its variability across clients, we discuss below the importance of the preheader in ensuring that emails are handled appropriately by their recipients. Given the small number of clients in widespread use, it is relatively easy for email senders to check the appearance of the preheader for the vast majority of users.



Facebook Inbox 634185 is your Facebook account recovery code - www.facebook.com/recover/code?u=...

Fig. 2. Message preheader example

- **Email header**. The header can only contain printable US-ASCII characters (except for carriage return, line feed, and colon), [7], and contains message metadata. It is broken into fields, each starting with a name followed by a colon (:). Examples of key header fields are briefly described below; many of these fields are displayed by the email client, along with the body of the message.
  - The **subject** field is chosen by the sender, and is commonly used to give a summary of the email.
  - The **to** and **from** fields contain the email addresses of the sender and recipient, as specified by the sender.
  - The **orig-date** indicates when the email was sent.
  - The **message-ID** is a globally unique email identifier, [8], [7].

  Email clients will typically display the *subject*, *to*, *from* and *orig-date* fields in lists of received emails; other header fields, e.g. the *message-ID*, are typically optionally viewable. For example, the Gmail *show original* button will reveal the entire email header.
- **Email body**. This constitutes the contents of the email. Email bodies were originally simply text strings, but MIME [9] allows bodies in a range of formats and with multiple parts. Today many email bodies contain an HTML-formatted part, which is by default what is displayed; this is thus the most important part for our analysis, and we look in detail at the information included in recovery email bodies below.
- **Email signature**. This is included at the end of the email body in a format determined by the sender's email client and the user. Its purpose is to provide information about the sender in an accessible form.

[2] http://hyperlinkcode.com/blog/html-johnson-box/

[3] The term 'preheader' is perhaps misleading since it is displayed after the subject header field, and actually occurs in the email after the header.

C. Email client features

A range of email information can be displayed by an email client, some of which is computed by the client rather than being contained in the email itself. Examples of information of this latter type include the following.

- **Sender Policy Framework (SPF)** [10] is an email-validation system designed to detect and block email spoofing. It allows receiving email exchangers to check that an incoming email from a domain comes from a host authorized by that domain's administrators.
- **DomainKeys Identified Mail (DKIM)** [11] aims to prevent email spoofing by allowing the receiver to check that an email claiming to have come from a domain was authorized by the owner of that domain.
- **Domain-based Message Authentication, Reporting and Conformance (DMARC)** [12] is an email-validation system designed to combat certain techniques often used in phishing and email spam, such as emails with forged sender addresses (in the *from* field) appearing to originate from legitimate organizations.

IV. Risks From Recovery Emails

We next review threats that can arise from poorly designed recovery emails. Of course, the use of email in password recovery has intrinsic risks (see, e.g., [4]), but we do not discuss them here. The list below was compiled by combining the discussion of design of the password recovery SMSs due to Gelernter et al. [5], with our own observations derived from our study.

- **Poor instructions**. It is important that the user is given clear instructions on how to use the email, and an indication of its sensitivity. Some emails give very brief instructions, e.g. limited to what the user should do to proceed, whereas others give almost too much information and are very hard to follow; this can be challenging, especially for non-technical users. Most seriously, the lack of clear instructions could lead to disclosure of the secret in the email, e.g. by forwarding the email to a third party, thus compromising password recovery.
- **Poor readability**. Some HTML-formatted emails use small fonts and/or small buttons, which can make reading them hard especially for those with impaired vision.
- **Lack of easily recognized source**. Some recovery emails lack clear information as to their source, e.g. a usable email address; in such cases the user may treat it as a phishing email or it may be blocked by a spam filter.
- **Email header or preheader leaking confidential information**. In some cases the temporary password or validation code is included in the header or preheader, which means that it might be available to anyone with temporary access to the phone, even if locked.
- **Lack of contact details**. Some emails lack contact details for use if the recovery process fails. Contact details are often necessary as the source email address may not be monitored for replies.
- **Lack of instructions if recovery not requested**. Recovery emails may lack instructions on what to do if password recovery was not requested, e.g. indicating an attempt to gain unauthorized access to the account.
- **Spam filter issues**. If the email is not constructed carefully, then there is an increased danger that it will be blocked by a spam filter [3], hence preventing the user from performing password recovery.

V. A study of existing practice

A. Scope of study

We examined 50 widely used English language sites. We looked at only 50 websites because we needed to manually register at each site, trigger a recovery email, and then carefully examine it. The average time to perform the first two steps was 20 minutes, and so it took around 17 hours simply to gather the data. The process could not be automated since user registration and password recovery requests require user interactions, e.g. solving a CAPTCHA or validating the email address by clicking on a link or entering a temporary code.

We chose the 50 highest-ranking sites employing email-based password recovery from the list of most-visited websites provided by majestic.com (chosen because it offers information free of charge). In a number of cases, we were not able to study a web site on the list; if so we simply moved on to the next site in the list, continuing until we had the results from 50 sites. Sites were omitted from our study if they did not offer the ability to create an account, if creating an account required taking out a subscription, or where an organisation (e.g. flickr.com and yahoo.com) operated multiple sites where we skipped second and subsequent related sites.

B. Methodology

We performed the following three steps with each of the 50 chosen websites. **Account creation/registration** involved giving a user name, password, and any other information requested. We **initiated password recovery** by making a password recovery request and recorded the received recovery email. Finally we **manually analysed** the password recovery email. In this analysis we first examined each email against the design risks listed in section IV, and in each case considered whether the email suffered from the risk. This generated the statistical results given in section VI-A below. We then looked at each of the four components of the email and noted any examples of particularly good or bad practice; the results of this examination are summarised in sections VI-B–VI-D.

VI. Results

A. Summary of risks

As described above, we start by summarising the observed frequency of the seven risk types given in section IV.

- **Poor instructions**. As many as 22 (44%) of the 50 emails contained no further instructions on the use of the provided link or code, as was the case for tumblr.com.

- **Poor readability**. Three (6%) of the emails used small fonts or condensed blocks of text, giving potential readability issues, e.g. www.ncbi.nlm.hih.gov.
- **Lack of easily recognised source**. In five (10%) of the cases, the *from* field contained an email address with no obvious link to the originating website making recognition problematic, e.g. dailymail.co.uk uses password@and.co.uk.
- **Email header or preheader leaking confidential information**. Only six of the 50 emails contained a secret code/temporary password; the others used a link or a combination of code and link; of these six, two (4%) leaked the code via the header or preheader field. These two websites are two of the most widely used sites, i.e. google.com leaked the code via the preheader and facebook.com included the code in the header.
- **Lack of contact details**. As many as 35 (70%) of the recovery emails did not provide any contact details, e.g. for use in the event of problems, although some of these (e.g. pinterest.com) provided links to web sites for further information (e.g. containing FAQs).
- **Lack of instructions if recovery was not requested**. 20 (40%) of the emails, including that from google.com, failed to provide any user guidance for the case where the recovery email was not requested.
- **Spam filter issues**. In our experiment, two (4%) of the emails were routed into a spam folder by the email client, e.g. the email from creativecommons.org. This suggests that the source of these emails could have done more to prevent such an undesirable event.

All 50 of the emails suffered from at least two of the above issues. This suggests that most websites need to improve the design of their recovery emails. Although some the identified issues seem minor, even if they cause a problem in just 1% of recovery attempts this could potentially result in a large number of compromised or unusable accounts.

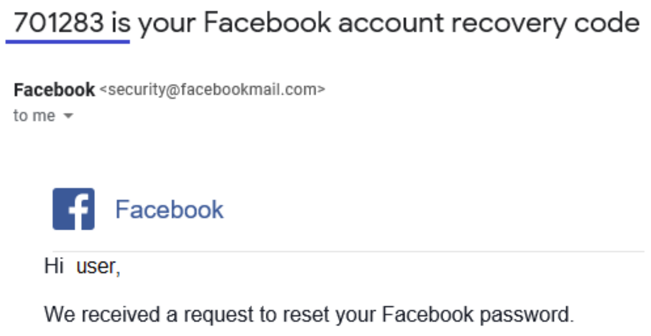
B. Email header

We focussed on the *subject* and *from* fields.

1) subject: Among the websites we tested, Facebook (see Figs. 2 and 3) was the only one which included a verification code in the *subject* field. An email client may display the *subject* fields of received emails even when a phone is locked, increasing the risk that the code will be seen by a third party.

2) from: This field includes an email address and, optionally, a 'name'.

- If the sender name is absent then the receiver might doubt its legitimacy. Six of the 50 emails did not state the name of the sender in this field, e.g. the www.ibm.com recovery email had *from* field 'ibmacct@us.ibm.com', i.e. without the name of the service. Similar issues arise in SMS-based password recovery [5].
- In some cases, the email address did not identify the type of service, i.e. relating to password recovery, e.g. no-reply@tumblr.com.
- Of the 50 recovery emails we examined, 21 (i.e. 42%) were sent from an email address to which replies are not permitted (e.g. noreply@bloomberg.com). Other websites did not use *noreply* but emails to the address bounced, e.g. washingtonpost.com. If a contact email is not provided then this is very unhelpful for users who encounter difficulties; it could also mean that the recovery email is blocked by an email client spam filter.
- Some emails were not constructed in such a way as to pass the SPF, DKIM and DMARC anti-spoofing tests, e.g. ibm.com failed DMARC testing.



701283 is your Facebook account recovery code

Facebook <security@facebookmail.com>
to me

Hi user,

We received a request to reset your Facebook password.

Fig. 3. Facebook password recovery email

C. Message body

28 of the emails contained a personalised **user greeting** (e.g. a name or email address), whereas others had a generic greeting or none at all. A personalised greeting is to be preferred since it helps reassure the user the email is genuine. Information as to when the request was initiated was included in three of the recovery emails we examined; this helps the user verify that the email corresponds to a legitimate request.

46 of the 50 emails included a URL link for password recovery. We have a number of observations.

- 41 (82%) of the emails did not specify how long the URL would remain valid. As discussed elsewhere [4], it is desirable for links to have a short lifetime, since they represent a security risk. Indeed, some tokens were sill valid even three months after the experiment, e.g. creativecommons.org and Bloomberg.com. In other cases the password recovery page was available after token expiry, allowing the user to generate a new token without user authentication, e.g. www.mozilla.org.
- 20 (40%) of the emails did not give a way of disabling the link in the event of accidental initiation of password recovery, increasing the risk of compromise.
- Many emails require the user to copy and paste the URL into a browser, which can expose the user to phishing attacks. E.g. an attacker could trigger password recovery for the user, and instruct the user to copy the URL in a recovery email into an attacker webpage.

Websites give varying advice for the case that the user did not request recovery. Some, e.g. www.dropbox.com and www.wix.com simply suggested ignoring the email; this may leave the user very concerned as to what is happening. Others suggest contacting the web service, e.g. by filling an online form, contacting a call centre (e.g. www.nytimes.com), replying to the email (e.g. www.vimeo.com), or clicking on a link (e.g. Facebook.com). Other sites, e.g. www.forbes.com, did not address the issue at all.

47 (94%) of the recovery emails failed to advise the user to protect the code or URL. Giving such a warning can help to reduce the risk of code sharing and phishing attacks. The same issue arises when using SMS-based password recovery.

D. Message signature

As many as 21 of the emails lacked a signature, e.g. www.wikipedia.org/. Other contained only the URL, e.g. www.microsoft.com and www.dropbox.com.

VII. Recommendations

The issues identified in our survey enable us to make the following recommendations for the design of recovery emails. We divide our recommendations under seven headings corresponding to the risks identified in section IV.

- **Instructions**. A personal greeting is desirable to increase user confidence. Where possible the email should include a clickable URL, and avoid asking the user to copy/paste the URL. Users should be warned against sharing or forwarding password recovery emails. More detailed information should be provided via links to avoid creating congested, unreadable emails. The recovery mechanism should have a limited validity period.
- **Readability**. The message should be easily comprehensible, i.e. the font/font size must be chosen with care.
- **Source recognition**. The name of the sender and the sender email address should match.
- **Email header and preheader design**. The header and preheader must not include any secret recovery information, e.g. a code or link, since preheaders can appear on a mobile phone screen even when locked. Users may also be tempted to perform recovery just by reading the preheader without fully viewing the email, thereby missing any instructions/warnings in the email. This issue can be more severe for users who view emails on a mobile phone, where the preheader can often be a 'notification'. Similarly, the *subject* field in an email header should state clearly the purpose of the email along with the service name.
- **Contact details**. The email should give a valid email address or a support link for use if the user is unable to recover a password. Ideally the email signature should include the name of the web service along with a logo, an email address, a phone number, a website URL and a physical address.
- **Instructions if recovery not requested**. Users should be given a means to disable the password recovery mechanism and to report unsolicited recovery emails to assure users of their security.
- **Spam filter issues**. The email should be generated and sent in such a way as to minimise the risk that it is blocked by a spam filter.

VIII. Concluding Remarks

We examined the password recovery emails sent by 50 of the most widely visited websites, and found that they all suffer from at least two types of defect. As a result it seems likely that some password recovery processes are likely to be unnecessarily prone to compromise. Although email-based password recovery allows the service provider to provide clear and detailed instructions, it would appear that some providers have given relatively little attention to the design of their recovery emails. In some cases the email structure, wording and content clearly give rise to potential vulnerabilities. Our survey has allowed us to make a series of recommendations regarding how best a password recovery email should be designed to maximise security and usability. There is clearly a need for further research in this important area, as well as new, more secure, ways of performing password recovery.

References

[1] S. Preibusch and J. Bonneau, "The password game: Negative externalities from weak password practices," in *Decision and Game Theory for Security — First International Conference, GameSec 2010, Berlin, Germany, November 22-23, 2010. Proceedings*, ser. Lecture Notes in Computer Science, T. Alpcan, L. Buttyán, and J. S. Baras, Eds., vol. 6442. Springer, 2010, pp. 192–207.

[2] J. Bonneau and S. Preibusch, "The password thicket: Technical and market failures in human authentication on the web," in *9th Annual Workshop on the Economics of Information Security, WEIS 2010, Harvard University, Cambridge, MA, USA, June 7-8, 2010*, 2010.

[3] S. L. Garfinkel, "Email–based identification and authentication: An alternative to PKI?" *IEEE Security & Privacy*, vol. 1, no. 6, pp. 20–26, 2003.

[4] F. A. Maqbali and C. J. Mitchell, "Web password recovery — a necessary evil?" *arXiv:1801.06730 [cs.CR]*, 2018, to be presented at FTC 2018.

[5] N. Gelernter, S. Kalma, B. Magnezi, and H. Porcilan, "The password reset MitM attack," in *2017 IEEE Symposium on Security and Privacy, SP 2017, San Jose, CA, USA, May 22-26, 2017*. IEEE Computer Society, 2017, pp. 251–267.

[6] D. Silver, S. Jana, D. Boneh, E. Y. Chen, and C. Jackson, "Password managers: Attacks and defenses," in *Proceedings of the 23rd USENIX Security Symposium, San Diego, CA, USA, August 20-22, 2014*, K. Fu and J. Jung, Eds. USENIX Association, 2014, pp. 449–464.

[7] P. Resnick (editor), *RFC 5322, Internet Message Format*, Internet Engineering Task Force, October 2008.

[8] S. Pasupatheeswaran, "Email 'message-ids' helpful for forensic analysis?" in *ADF 2008 — 6th Australian Digital Forensics Conference, Perth, Australia, 3 December 2008*. School of Computer and Information Science, Edith Cowan University, Perth, Western Australia, 2008.

[9] N. Freed and N. Borenstein, *RFC 2045, Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies*, Internet Engineering Task Force, November 1996.

[10] S. Kitterman, *RFC 7208, Sender Policy Framework (SPF) for Authorizing Use of Domains in Email, Version 1*, Internet Engineering Task Force, April 2014.

[11] D. Crocker, T. Hansen, and e. M. Kucherawy, *RFC 6376, DomainKeys Identified Mail (DKIM) Signatures*, Internet Engineering Task Force, September 2011.

[12] M. Kucherawy and e. E. Zwicky, *RFC 7489, Domain-based Message Authentication, Reporting, and Conformance (DMARC)*, Internet Engineering Task Force, March 2015.