It's Not What You Know, But Who You Know
A social approach to last-resort authentication

Stuart Schechter
Microsoft Research
StuS@microsoft.com

Serge Egelman
Microsoft Research
Egelman@cs.cmu.edu

Robert W. Reeder
Microsoft (TUX)
RoReeder@microsoft.com

ABSTRACT

Backup authentication mechanisms help users who have forgotten their passwords regain access to their accounts—or at least try. The security and reliability of today's backup authentication mechanisms have significant room for improvement. We designed, built, and tested a new authentication system that employs *social-authentication*. The system employs *trustees* previously appointed by the account holder to verify the account holder's identity. We ran three experiments to determine whether the system could (1) reliably authenticate account holders, (2) resist email attacks that target trustees by impersonating account holders, and (3) resist phone-based attacks from individuals close to account holders. Results were encouraging: seventeen of the nineteen participants who made the effort to call trustees authenticated successfully. However, we also found that users must be reminded of who their trustees are. While email-based attacks were largely unsuccessful, stronger countermeasures will be required to counter highly-personalized phone-based attacks.

Author Keywords

Security, Privacy, Usability Testing and Evaluation

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: Miscellaneous

INTRODUCTION

As long as websites authenticate users via credentials that are either memorized (e.g. passwords) or stored (e.g. smartcards), users will inevitably forget or lose them.

The 'secret' personal questions and alternate email addresses currently used for backup authentication by webmail providers are unfortunately unreliable. For personal questions, prior and concurrent research has shown that users forget their answers and their acquaintances may be able to guess them [14, 9, 12]. An account holder who tries to authenticate via an alternate email address may find that the configured address expired upon a change of job, school, or Internet service provider. As other websites rely on email addresses to authenticate their account holders when passwords fail, it is especially important for us and other webmail providers to have a secure and reliable authentication mechanism of last resort.

We thus designed and built a new backup authentication mechanism of last resort and studied its performance by making it appear to be part of Windows Live ID. The mechanism uses *social authentication*, in which account holders initially appoint and later rely on account *trustees* to help them authenticate. To regain access to their accounts, account holders contact their trustees by phone or in person, so that their trustees may recognize them by their appearance or voice. A trustee who recognizes an account holder may provide him or her with an *account-recovery code*. An account holder must present a sufficient number of these codes (e.g. two codes from any of four possible trustees) to authenticate.

The overall success of an authentication mechanism depends on four important measurement categories:

Setup and maintenance costs:
The time or effort required of the account holder to configure or reconfigure the authentication mechanism

Efficiency:
The time or effort required of the account holder each time he or she authenticates to the system

Reliability:
The likelihood that the account holder can successfully authenticate his or her identity

Security:
The time or effort required to impersonate (falsely authenticate as) an account holder, or likelihood of doing so successfully.

Reliability is especially important for a backup authentication mechanism of last resort: account holders who find themselves needing to use this mechanism may have no other chance to regain access to their accounts. Yet, reliability cannot be achieved at the expense of security; if a backup authentication mechanism is less secure than the primary mechanism it supports, its very existence will make users' accounts less secure. Fortunately, backup authentication occurs less often than primary authentication, and so efficiency may be sacrificed to achieve reliability and security.

CHI 2009, April 4-9, 2009, Boston, Massachusetts, USA.

We conducted three experiments to test our new social authentication mechanism. In the first, we asked Hotmail users to authenticate using this mechanism and measured its reliability and efficiency. The two remaining experiments tested the system's security against attacks in which someone other than the account holder requests account-recovery codes from the account holder's trustees. In our second experiment, trustees received a form email from a new email address that had been opened in the account holder's name. In our third experiment, trustees received a call from a close friend or family member of the account holder. The email requests gauged the system's vulnerability to automated (scalable) attacks whereas the phone requests gauged performance under conditions made extremely favorable to an attacker.

Our experiments, the first ever performed on a social authentication system, were designed with an emphasis on ecological validity. We wanted trustees to believe that revealing an account-recovery code to the wrong person could actually result in the compromise of the account they were entrusted to protect. Thus, trustees were not informed they were participating in a study when they encountered our simulated attacks. We deployed our prototype and made it accessible to the public at `recover.live.com`, where it was made to appear as a fully operational feature of Windows Live ID that could indeed be used to reset account holders' passwords.

BACKGROUND AND RELATED WORK

To place our work in context, it is important to first understand the limitations of passwords as a primary authentication mechanism and the efficacy of existing backup authentication mechanisms.

Primary authentication failure: forgotten passwords

Password memorability was recently studied in the laboratory by Vu et al. in 2007. After one week, 12.5% of participants had forgotten their six-character alphanumeric passwords. Of the participants who had to remember five account passwords, 25% of them forgot at least one [13]. A 2004 survey by SafeNet found that 47% of respondents forgot their passwords and needed to request at least one password reset annually [11].

Some password mismatches result when users mistype passwords or cannot remember which of their passwords to use. Brostoff and Sasse observed that allowing as many as ten chances to enter a password would reduce password reset requests [3]. Security practitioners have already put these findings into practice. For example, Windows Live ID already gives users ten password-entry opportunities and gives even more after requiring the user to solve a CAPTCHA.

Modern web browsers have integrated password managers that remember and enter users' passwords for them. Those who use these features need not enter their passwords as often, and thus may be less likely to remember their passwords when they do need to enter them. These users may resort to backup authentication if they lose the data in their password managers, replace their computers, or start working from new computers.

Backup authentication using personal questions

The use of personal questions for authentication was studied by Zviran and Haga in 1990 [14]. They suggested using either fact-based (e.g. "what university did you attend?") or opinion-based (e.g. "what is your favorite color?") questions as an alternative to password-based authentication. They found that participants remembered 78% of their answers, but that significant others were able to guess 33% of their answers. Podd et al. conducted a similar study in 1996, and found similar recall rates (80%) and higher guessing rates (39.5%) [9]. In 2008, Rabkin conducted a study of twenty bank websites that use personal questions as a backup authentication mechanism. He found that many of the questions were either not applicable to over 15% of the general public, not memorable, ambiguous, easily guessable with no knowledge of the victim, or easily guessable with minimal knowledge of the victim [10].

Jakobsson et al. proposed a question-based backup authentication scheme that relies on preference-based questions from online dating websites [7]. Responses are provided on a preference scale as compared to the free response answers in earlier schemes. By employing a large number of questions (e.g. 16) to be configured initially and answered during authentication, and not requiring all answers to be correct, the researchers were able to achieve low false rejection rates. Traditional schemes using free response questions could also benefit from having multiple questions; many banking websites already require account holders to answer a subset of three or more questions.

Many personal questions can be guessed with only limited knowledge of the victim. For example, in 2008 the answer to the question "Where did you meet your spouse?" was implicated in the compromise of a Yahoo! account belonging to then vice-presidential candidate Sarah Palin [2] (the answer was "Wasilla High School").

In parallel with this study, we conducted a separate study of both the guessability and memorability of the personal questions used by the top four webmail providers [12]: AOL, Google, Microsoft, and Yahoo. We invited 65 pairs of previously-acquainted individuals (parners) to answer these questions, then offered them an incentive to try to guess their partners' answers. Unlike previous studies, which only focused on significant others [14, 9], we sought less intimate pairs by recruiting friends and coworkers. To gauge trust, we asked participants whether they would lend their study partner their Hotmail passwords. Though participants who were close enough with their partners to be trusted with their password guessed more answers than those who weren't, the difference was smaller than a factor of two. We also measured recall rates after four to six months and found numbers similar to those of Zviran et al. and Podd et al.

Alternative backup-authentication mechanisms

The ubiquity of mobile phones has made them an attractive option for backup authentication. Banks, such as by Australia's CommonwealthBank [4], already send SMS messages containing authorization codes to supplement primary authentication for high-risk transactions. However, authenticating users by their mobile phone alone is risky as phones are frequently shared or lost—an estimated $60,000$ are lost each year in New York City cabs alone[5].

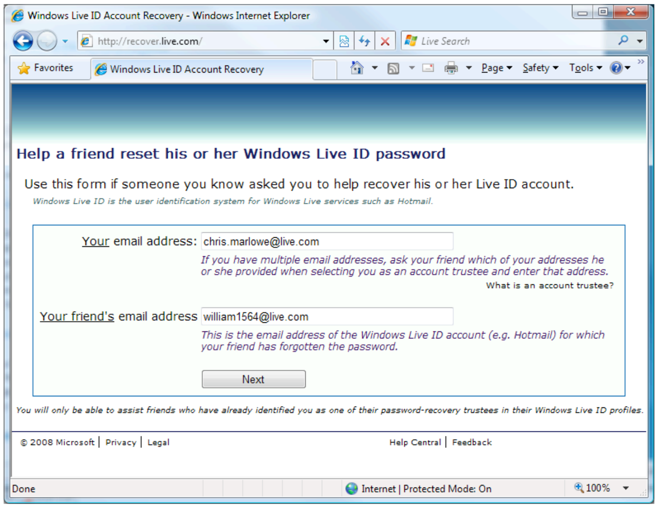
Some websites offer last-resort authentication through their customer-support departments. However, introducing human customer support teams may not provide a strong advantage over automated systems, as information used by support staff to authenticate an account holder may be no better than the information available to the automated systems. For example, Google's technical support form for password recovery captures the IP address from which requests take place. Users are asked for their account creation date, last login date, and the last passwords they remember [6]. Alas, last login dates have a tight distribution (most are recent) and only the newest of users are likely to remember their account creation date. Microsoft's form asks for the names of Hotmail folders and contacts [8]. This information could be read over users' shoulders. Furthermore, many users are unaware this information is used for authentication and would thus not know to withhold it should anyone ask.

Trustee-based authentication

The concept of shifting the responsibility to authenticate an individual from one party to another is not new. Authenticating users via an alternate email address shifts the responsibility to authenticate to the provider of that alternate address. In organizations, the responsibility to authenticate a user who fails primary authentication is often shifted to system administrators, corporate security, or other support staff. Microsoft has long employed a form of trustee-based account recovery for its own employees: if an employee forgets her account credentials, her manager or coworkers can request a temporary password on her behalf.

In 2006, Brainard et al. of RSA proposed a two-factor primary authentication system (PIN and token) for enterprise use in which a user who lost her token could receive help from a pre-selected trustee they called a "helper" [1]. In their system, the trustee authenticates using her two factors in order to generate a "vouchcode" that substitutes for the account holder's lost token. To our knowledge, no usability results have been made available.

Whereas RSA's system is designed for primary authentication, ours is designed for last-resort authentication. We cannot assume that our users can contact a system administrator when all else fails. Whereas RSA's system requires users to select a helper (trustee) who has an account on the same system, ours requires only that trustees have email addresses.


Windows Live ID Account Recovery - Windows Internet Explorer
http://recover.live.com/
Live Search
Favorites
Windows Live ID Account Recovery
Page
Safety
Tools
Next
© 2008 Microsoft | Privacy | Legal
Help Central | Feedback
Done
Internet | Protected Mode: On
100%

Help a friend reset his or her Windows Live ID password

Use this form if someone you know asked you to help recover his or her Live ID account.

Windows Live ID is the user identification system for Windows Live services such as Hotmail.

Your email address: chris.marlowe@live.com

If you have multiple email addresses, ask your friend which of your addresses he or she provided when selecting you as an account trustee and enter that address.

What is an account trustee?

Your friend's email address william1564@live.com

This is the email address of the Windows Live ID account (e.g. Hotmail) for which your friend has forgotten the password.

You will only be able to assist friends who have already identified you as one of their password-recovery trustees in their Windows Live ID profiles.

Figure 1. Initiation. Trustees enter their email address and the address of the account holder.

ACCOUNT RECOVERY VIA SOCIAL AUTHENTICATION

We designed, built, and deployed an account recovery (password reset) mechanism employing social authentication, in which users could authenticate by obtaining account-recovery codes from three of four previously selected trustees. We deployed the system at `recover.live.com` where it was made to appear as a fully functional feature—though members of the public could not sign up to use the system for their own accounts.

The primary threat to a social authentication system is that an attacker – someone other than the account holder – will convince or trick the account holder's trustees to vouch that the attacker is the account holder. That is, the attacker would request and receive the information required to obtain an account-recovery code. The attacker might do this by impersonating the victim or by convincing the trustee that he or she is acting on behalf of the victim.

In this section we provide an overview of the system, user experience, and countermeasures to defend against attacks.

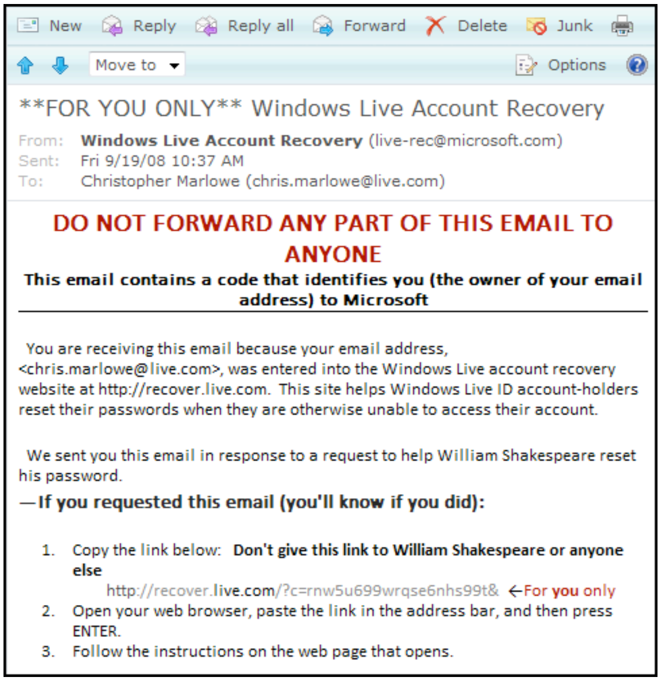
Configuration

Our social authentication mechanism required that users provide the names and email addresses of four trustees in advance of use. We did not test the configuration step as part of this study. We also chose not to inform trustees of their selection: we feared this might lead them to ask account holders about the system, learn it was a prototype, and thereby change their security behavior.

Recovery

When an account holder needs to recover his[1] account, he must obtain account-recovery codes from his trustees. Account holders instruct their trustees to visit the account-recovery system at `recover.live.com`. We encourage account holders to call or visit their trustees in person. Because we

[1] For clarity, we use masculine pronouns for the account holder and feminine pronouns for trustees.



FOR YOU ONLY Windows Live Account Recovery

From: **Windows Live Account Recovery** (live-rec@microsoft.com)
Sent: Fri 9/19/08 10:37 AM
To: Christopher Marlowe (chris.marlowe@live.com)

DO NOT FORWARD ANY PART OF THIS EMAIL TO ANYONE

This email contains a code that identifies you (the owner of your email address) to Microsoft

You are receiving this email because your email address, <chris.marlowe@live.com>, was entered into the Windows Live account recovery website at http://recover.live.com. This site helps Windows Live ID account-holders reset their passwords when they are otherwise unable to access their account.

We sent you this email in response to a request to help William Shakespeare reset his password.

—If you requested this email (you'll know if you did):

1. Copy the link below: **Don't give this link to William Shakespeare or anyone else**
   http://recover.live.com/?c=rnw5u699wrqse6nhs99t& ←For you only
2. Open your web browser, paste the link in the address bar, and then press ENTER.
3. Follow the instructions on the web page that opens.

Figure 2. Trustee-authentication email. This email contains a link that identifies the trustee to our website.

discourage trustees from responding to requests for account-recovery codes that arrive via email or text messages (they are easy to spoof), we also discourage account holders from contacting their trustees using these channels.

We were not sure how many account-recovery codes should be required to authenticate an account holder. We configured the system to require a threshold of three codes so that we could measure the time required to obtain both the second and third code. To obtain an account-recovery code, a trustee must perform four steps.

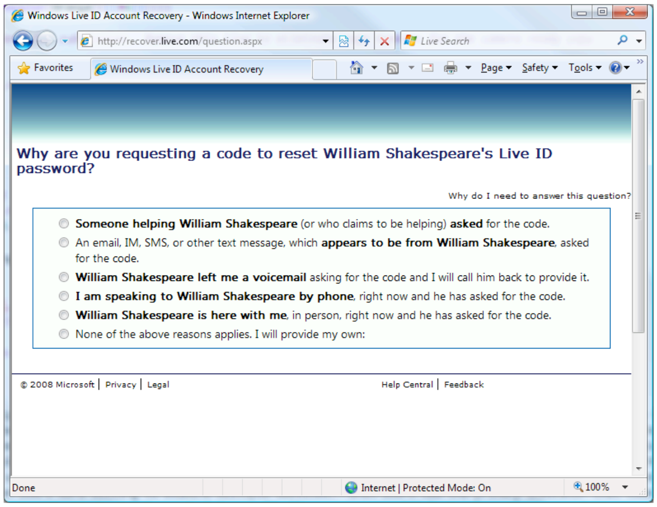
Initiation

When the trustee first visits the account recovery system, she is asked to enter her email address and the address of the account holder she is assisting (Figure 1).

Trustee-authentication email

Next, the trustee receives an email from the account recovery system (Figure 2). If she is indeed a trustee for the specified account holder, the system creates a record to track the request and the email sent to the trustee will contain a code pointing to this record. The trustee copies this link into her browser's address bar to continue.

This emailed link and code are all that are required to prove the trustee's identity and retrieve the account-recovery code. An attacker who could convince a trustee to forward the email would be able to retrieve the code. Two countermeasures against this attack are the email's subject, which begins with "**FOR YOU ONLY**", and the message body, which begins with a conspicuous warning "do not forward any part of this email to anyone" (see Figure 2).


Windows Live ID Account Recovery - Windows Internet Explorer
http://recover.live.com/question.aspx
Live Search
Favorites
Windows Live ID Account Recovery
Page
Safety
Tools

Why are you requesting a code to reset William Shakespeare's Live ID password?

Why do I need to answer this question?

○ **Someone helping William Shakespeare** (or who claims to be helping) **asked** for the code.
○ An email, IM, SMS, or other text message, which **appears to be from William Shakespeare**, asked for the code.
○ **William Shakespeare left me a voicemail** asking for the code and I will call him back to provide it.
○ **I am speaking to William Shakespeare by phone**, right now and he has asked for the code.
○ **William Shakespeare is here with me**, in person, right now and he has asked for the code.
○ None of the above reasons applies. I will provide my own:

© 2008 Microsoft | Privacy | Legal Help Central | Feedback

Figure 3. Query of intent. The trustee is asked to report why she is requesting an account-recovery code.

Query of intent

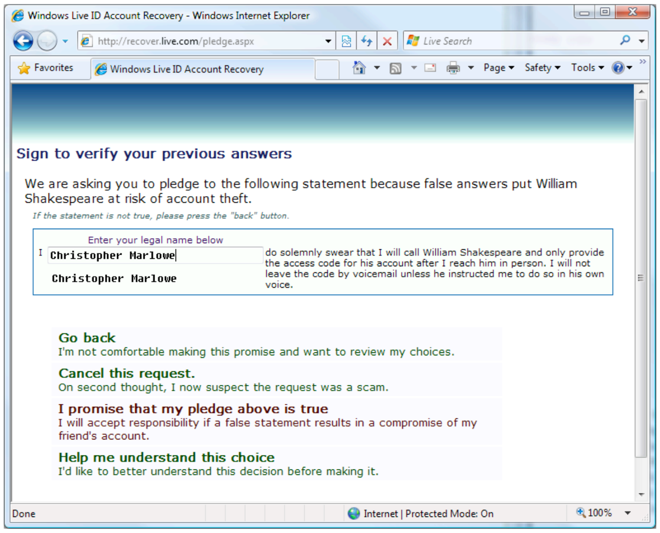
When the trustee pastes the link from the trustee-authentication email into her browser, she is asked to explain why she is requesting an account-recovery code by choosing from a set of options, illustrated in Figure 3. These options may convey that she has heard from the account holder personally or that she is responding to a request from a third party.

The options that indicate the highest risk of fraud are listed at the top in order to maximize the chance that the trustee will read them before making a choice. If the trustee chooses either of the top two options, she encounters a warning page that describes telltale signs of fraud and encourages her to contact the account holder by phone or in person. She is, however, given the option to disregard these warnings and continue.

Pledge

Finally, the trustee is asked to pledge to her previous answer and to her understanding of the potential consequences of giving an account-recovery code to someone other than the account holder. This pledge requires her to type her name, as provided by the account holder, and to press a button that says "I promise the above pledge is true". For example, if a trustee reports receiving a request from the account holder via voicemail, she would be asked to pledge that she will only provide a code after she reaches him "in person", as illustrated in Figure 4.

After the trustee has signed the pledge, the system presents the six character account-recovery code. If this is the first account-recovery code requested for this account holder, the system will then email the remaining trustees to notify them of the event and encourage them to call the account holder. To further protect against attack, the account holder will be notified whenever he next logs in (or if he is already online). If an attack were underway, a call from his trustees would alert the account holder to login and halt the recovery process before the attacker can complete it.


Windows Live ID Account Recovery - Windows Internet Explorer
http://recover.live.com/pledge.aspx
Live Search
Favorites
Windows Live ID Account Recovery
Page
Safety
Tools
Done
Internet | Protected Mode: On
100%

Sign to verify your previous answers

We are asking you to pledge to the following statement because false answers put William Shakespeare at risk of account theft.

If the statement is not true, please press the "back" button.

Enter your legal name below

I Christopher Marlowe do solemnly swear that I will call William Shakespeare and only provide the access code for his account after I reach him in person. I will not leave the code by voicemail unless he instructed me to do so in his own voice.

Christopher Marlowe

Go back
I'm not comfortable making this promise and want to review my choices.

Cancel this request.
On second thought, I now suspect the request was a scam.

I promise that my pledge above is true
I will accept responsibility if a false statement results in a compromise of my friend's account.

Help me understand this choice
I'd like to better understand this decision before making it.

Figure 4. Pledge. The trustee must confirm previous statements and agree on a course of action.

EXPERIMENT 1: RELIABILITY

In our first experiment, we tested the reliability and efficiency of our social-authentication mechanism. Reliability is the fraction of users who could successfully authenticate using social authentication and efficiency is measured as the time required to do so (effort is more difficult to quantify).

Design

During a previous study of our Hotmail users, which took place from March to May of 2008, we introduced the concept of social authentication and requested that participants consider who they would choose as trustees for their Windows Live ID account. We asked each participant to list four trustees' names, email addresses, and the relationship between the participant and the trustees.

In September, we invited 43 of our prior participants to earn a software gratuity by obtaining account-recovery codes from three of the four trustees they had previously identified. Emails inviting participants to the study were sent at evenly divided intervals[2] over a week. This distribution ensured that we would not favor a particular time of the day or week that might be better or worse for contacting trustees.

Most users locked out of an online account will act with some urgency to regain access. To induce a similar sense of urgency, we ran a contest based on participants' relative performance: the fastest quartile received an Amazon.com gift card worth $50, the second quartile received one worth $25, the third quartile $10, and the last quartile received no gift card. All participants who completed the task and emailed a final survey within one week received a software gratuity.

Participants were not initially given the names and email addresses of their trustees. Those who had since forgotten them could choose to request them via a web page that assessed a two-hour penalty against their contest time.

Results

The 43 individuals we invited to participate in this experiment are categorized in Table 1. Row 1 shows that 13 of those invited never loaded the webpage containing the instructions for this experiment. They may have not received the email or decided not to open the instructions.

Of the 30 remaining invitees, four never responded to either our initial survey request or a follow-up request. The follow-up request offered a $10 Amazon.com gift card for simply replying to an email asking if they had tried to start the task or if they encountered any problems.

Of the 26 remaining invitees, two reported that they could not remember their trustees. The study instructions provided a link to a webpage that listed participants' trustees for use in completing the task. We responded with an email pointing this out, but received no response. Two others, who replied to our $10 offer, informed us that they were too busy to start the task in the first place.

Of the 22 remaining, one opted out because she did not want to contact her trustees citing fears that "trustees would think the message could contain unwanted viruses". Two others never found time for the task: One left a voicemail for the trustee she wanted to start with, but did not find a time to talk during the one week deadline; one replied that the instructions were complicated and that the task would require "an evening to call everyone." Neither reached a single trustee.

Of the 19 remaining, one participant had initially entered invalid email addresses for her trustees, and so they were unable to authenticate themselves in order to help her. This left 18 participants: 17 obtained three account-recovery codes and one participant obtained two.

The participant who was unable to obtain the third code reported that two of her trustees were unwilling to provide them. When asked why, she reported that she had initially requested these codes by sending SMS messages to her trustees. As SMS messages are easily forged, the system instructs trustees not to provide account-recovery codes under such circumstances.

One disappointing result was that 13 of the 18 participants who obtained two or more account-recovery codes (72%) required the system to provide them with a list of their trustees' names and email addresses. Thus, it appears inevitable that the system require some hints to remind users of the identity of their trustees. These hints might also allow attackers and others to identify an account holder's trustees.

Figure 5 illustrates the time participants required to obtain codes and complete the account-recovery task. Of those 17 who obtained three account-recovery codes, 8 (47%) did so in under 100 minutes and 12 (66%) in under a day. All but three of the 17 (88%) did so in two days. One of the remaining three obtained all three recovery codes on her fifth day, from which one might infer that she waited between reading the instructions and starting the task.

[2] A small number of variations resulted from a bug.

row	grouping											and then there were...
	43	*prospective participants were sent invitiaton emails*										*43*
1	13	never loaded the page containing instructions for this task									*30*	30%
2	4	never responded to any of our survey requests								*26*	13%	9%
3	2	couldn't remember trustees & didn't look them up							*24*	8%	7%	5%
4	2	too busy to attempt task						*22*	8%	8%	7%	5%
5	1	worried trustees would fear viruses in emails					*21*	5%	4%	4%	3%	2%
6	2	never spoke with any trustees				*19*	10%	9%	8%	8%	7%	5%
7	1	configured invalid emails for all trustees			*18*	5%	5%	5%	4%	4%	3%	2%
8	1	only successful with 2 trustees		*17*	6%	5%	5%	5%	4%	4%	3%	2%
9	17	successful with 3 trustees	*0*	100%	94%	89%	81%	77%	71%	65%	57%	40%

Table 1. Experiment 1 Invitees.
Each group is presented as a number of individuals followed by a description of their outcome. The italicized integer that follows is the number of participants remaining once this group, and all above it, have been removed from the set of 43 individuals invited to the study. Below each of these integer column headings, the remaining participants are broken down by the percentage that fall into each of the remaining groupings.
For example, the second column from the right breaks down the group of invitees who had loaded the task instructions (the column heading is in the row that excludes those who did not load the instructions). The bottom-most entry in that column is the percentage (57%) of that group who successfully obtained codes from three trustees (as indicated by the grouping description for the bottom row).
Because of rounding, not all columns add to 100%.

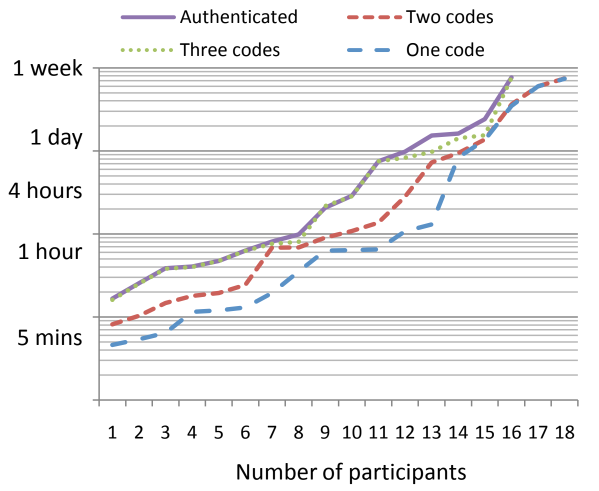


Figure 5. Distribution of time to complete each step of Experiment 1

Analysis

Despite the rough edges in our first implementation of social-authentication, seventeen of the nineteen participants who made the effort to reach one trustee (89%) were able to authenticate. In contrast, the highest aggregate recall rate for personal questions in prior and concurrent work is 80% [9, 12]. Furthermore, a real-world social-authentication system would likely reduce the threshold of required account-recovery codes for dormant accounts (e.g. those that had not been logged into successfully within the past week). For a two-trustee threshold, 18 of the 19 participants (94%) who made the effort to contact at least one of their trustees would have eventually regained access to their accounts.

Recall that our system did not alert trustees of their selection by account holders or of the implications of that selection. If we had alerted trustees and required them to opt-in, the one participant who entered invalid addresses for all her trustees would have detected this by the end of the opt-in period.

Our greatest remaining design challenge for ensuring reliability is to find a way to remind account holders who their trustees are. Our system could provide hints, such as portions of names and email addresses. Fake addresses the account holder would not recognize could even be mixed in to confound attackers. Trustees could be made privy to each others' identities so that if the account holder could reach one trustee, she could rely on that trustee to remind her who her other trustees were. Yet another option would be to move the burden of initiating contact from the account holder to the trustee, who would be notified whenever a password recovery was requested and asked to contact the account holder.

EXPERIMENT 2: SECURITY AGAINST FORGED EMAIL

Perhaps the easiest way to attack a large number of accounts at once is to automatically create webmail accounts in victims' names and email requests for account-recovery codes to those they associate with. We designed the following experiment specifically to test the security of our social authentication mechanism against such forged-email attacks.

Design

As we had invested significant time in building our social authentication system and safeguarding it against attack, we were concerned about our ability to remain unbiased in attempting to create the best possible message to use in this attack. To counter any such bias, we recruited assistance in crafting and editing these emails from Dan Kaminsky, a security researcher from the hacker community renowned for his penchant and talent for breaking systems. As importantly, Mr. Kaminsky was not previously acquainted with any of our researchers, met with us in person only once, and was in no way invested in the success of our system in repelling attacks. Two generic messages resulted from our brief collaboration. The first asked the trustee to generate an account-recovery code and respond with that code:

> I'm really sorry to bother you, but I'm having problems getting to my usual Hotmail account – [partici-

> pant's webmail addresss] (my password doesn't work) and can't remember the answer to my secret question. Can you help? I gave your name and email address as someone who could help get me get my account back. I could really use your help!
>
> The website for getting the code is recover.live.com. All you have to do is go there, follow the instructions, say that you're talking to me by phone (this option gets you through the fastest), and send back the code as soon as you can.
>
> Thanks so much!

The other message differed only in the second paragraph, which encouraged the trustee to forward the email our system sends to authenticate the trustee. If the trustee did so, the participant would then have the link required to generate an account-recovery code for that trustee.

> The website for getting the code is recover.live.com– it's a Microsoft site and a real pain to use. I've already filled out the web form for you so you don't have to use it, just toss me a copy of the email they send you. Look for it to come from "Windows Live Recovery Service". All you have to do is send it.

To measure the efficacy of attacks using these emails, we invited fifteen Hotmail users to our laboratory for two-hour sessions. We briefed participants on how social authentication works and demonstrated how they could use our system to recover a Windows Live ID account. We then asked participants to list up to fifteen individuals who they would consider using as trustees, in order of preference; those deemed most suitable by the participants were listed first. We encouraged participants to use their mobile phones, email accounts, and other online resources to locate email addresses of potential trustees, so long as they did not contact them. Participants provided each trustee's name, email address, the nature of their relationship, and indicated whether the trustee knew that the participant was at a research study. To assuage participants' privacy concerns, we informed them that they would not be required to give us their trustee lists and that we would not ask them to decide whether to give us their lists until after we briefed them on how we would use the information.

After participants had selected their trustees, we briefed them on the purpose of the study and asked them to create a new Hotmail account in their own name. They selected the email address for their new account by typing their full name, minus spaces, as the proposed username. If this address was not available, they used the first alternative proposed automatically by Windows Live ID's account-creation system.

We then asked participants to email each of their trustees using one of the two form messages, using a coin flip to select which message went to each trustee. Participants sending the second message would first fill out the form needed to send the trustee a trustee-authentication email and then send the message that requested they forward this email.

Trustees thus received an email from a newly created account that happened to belong to our laboratory participant, but could have just as easily been created by someone impersonating our laboratory participant who knew only her name, email address, and the email address of the trustee. If the account had been created by someone impersonating our laboratory participant, a trustee who replied with an account-recovery code would be putting the laboratory participant's account at risk. (Unbeknownst to the trustees, the code could not actually be used to reset the laboratory participant's password.)

Participants received a gratuity worth \$2 for each trustee they emailed and \$5 for each trustee who fell for the attack. Participants were forbidden from further initiating any other contact with their trustees or others. For the remainder of the session participants remained sequestered and we monitored their incoming mobile phone calls, their newly created 'attack' email accounts, and their existing account for contact from their trustees. When participants received calls from their trustees, the trustees were encouraged not to discuss the experiment with others.

We recorded emails containing account-recovery codes or forwarded emails that were sent to the new 'attack' email accounts as successful attacks. If a trustee called or sent email to the participant's real account we marked the attacks as failures. After the conclusion of the study, we followed up with participants to ask them to self-report any additional contacts by their trustees. We offered a fixed gratuity (\$10) for responding so as to remove any incentive to misreport outcomes.

Results

The fifteen participants identified and emailed a total of 118 trustees, 21 of whom were discarded because they knew the participant was currently at the study. This left 97 trustees, with each participant contacting a median of six ($\mu = 6.467$, $\sigma = 4.138$).

Participants spent much of the first hour filling out forms and learning the system, leaving roughly an hour window for us to observe trustees' responses while participants remained sequestered. Incoming phone calls, SMS messages, and emails provided evidence that trustees had encountered the email requests. We had no way of detecting if other trustees had encountered the requests and chosen to ignore them.

Only a total of sixteen of the trustees (16% of 97) responded during the sequestration period, two of whom (12.5%) sent codes to the 'attack' account. We observed no significant differences in response to the two different attack messages, so we analyzed all the results as a single data set. Tables 2 and 3 summarize trustees' responses.

In follow-up surveys we asked which trustees had since sent account-recovery codes to the new 'attack' email account and which contacted the participant about the email via a more trusted channel (by phone, a known-valid email ad-

	while participants sequestered in lab		*after lab session (via survey)*		*total*	
no response	-		33	(45%)	33	(37%)
success	2	(12%)	2	(3%)	4	(4%)
failure	14	(88%)	38	(52%)	52	(58%)
Total	16		73		89	

Table 2. Outcomes of "attack" emails soliciting account-recovery codes from trustees. Success indicates a response containing either an account-recovery code or enough data to generate a code.

	while ppts in lab	*after session (via survey)*	*total*
by phone	10	25	35
via *genuine* email address	3	14	17
by SMS (text) message	1	-	1
in person	-	9	9

Table 3. Reasons "attack" emails were deemed failures. Emails were deemed failures if trustees did not respond, but instead contacted the account owner (our laboratory participant) through another channel. Several participants' trustees contacted them through multiple channels, so column totals exceed the row labeled *failure* in Table 2.

dress belonging the the account holder, SMS, or in person). This expanded our data set by an additional 73 trustees to a total of 89. Two additional trustees sent account-recovery codes to the 'attack' account–though one became concerned and called the victim after sending it. Another 38 alerted the account holder about the emails. We received no data on 33 of the remaining trustees (37% of 89) and assume they must have either not noticed or ignored the attack email. We combine the sequestered data with the self-reported post-survey data in Tables 2 and 3. Note that trustee responses cannot be treated as independent trials, as multiple trustees shared the same study participant.

To understand the implications of these figures, it's worth looking at them from the attacker's perspective. An attacker's email has three possible outcomes, we label **c**, **d**, and **r**.

c: the attacker receives an account-recovery code
d: the attacker's email is dropped or ignored
r: the trustee reports the attack to the account holder.

Assuming that account holders will be able to halt the recovery process if alerted to an attack, the attackers will need to reach the threshold before any trustees report the attack. Each of the following sequences represent orderings that result in a successful attack against a threshold of three codes:

ccc, **dccc**, **cdcc**, **ccdc**

Each of the following sequences result in a successful attack against a threshold of two codes:

cc, **cdc**, **dcc**, **cddc**, **ddcc**, **dcdc**

The probabilities of successful attack sequences are thus:

$$P_c^2\left(1+2P_d+3P_d^2\right) \quad [\textit{for threshold of } 2]$$
$$P_c^3\left(1+3P_d\right) \quad [\textit{for threshold of } 3]$$

For example, suppose the figures in the rightmost column of Table 2 reflect the general population, despite the small sample size of trials. We can set $P_c = \frac{4}{89}$ and $P_d = \frac{33}{89}$ and calculate the probability of a successful attack to be 0.46% for two account-recovery codes and 0.019% for three.[3]

Analysis

The example above illustrates the benefits of requiring three account-recovery codes initially, and only later relaxing the threshold to two. While we cannot rule out the possibility that a clever attacker can devise an attack email far superior to the one we used in this study, these initial results are promising.

Furthermore, there are reasons to believe the system may be more secure than the above equations imply. For one, the equations do not take into account the notification sent by the system to the account holder whenever a trustee obtains an account-recovery code. If the account holder accesses his account after the first account-recovery code has been obtained, but before the threshold is reached, he can be notified of the account-recovery process underway and given the option to halt it.

Finally, it is important to note that the equations above presume that the attacker knows who the account holder has selected as trustees. If the attacker emails a larger set of potential trustees, those who are not trustees may still report the incident to the account holder. More importantly, the system may detect trustee-authentication requests that have invalid trustee/account holder relationships as indicators of a potential attack on that account holder.

EXPERIMENT 3: SECURITY AGAINST PHONE REQUESTS

In our final experiment, we wanted to determine how easily someone already acquainted with an account holder could convince the account holder's trustees to reveal an account-recovery code or trustee-authentication email.

Design

We recruited 9 pairs of participants: three pairs of spouses or spousal-equivalents, one son and mother in-law, four pairs of friends of five or more years, and one pair of friends of one year.

We asked participants to identify 10 potential trustees, asking for the trustees' phone numbers in addition to the information provided by participants in Experiment 2.

After both of the partners had identified their trustees, we revealed the purpose of the study: participants would take turns calling their partners' trustees and attempt to retrieve either a valid account-recovery code or trustee-authentication email for their partner's account.

Each time a participant (the *caller*) connected to one of the other's trustees, both participants received a $2 bonus. If

[3] We cannot calculate confidence intervals because the individual trials were not independent events.

the caller succeeded in either getting an account-recovery code or a trustee-authentication email out of their partner's (the *target*'s) trustee, both participants received a success bonus of $5. By providing the target participant of a call the same bonus received by the calling participant, we hoped to make the targets more willing allow callers to contact their trustees. We also hoped to reduce any hard feelings if the caller was successful. Outside of allowing calls to be made, the target participant on a call was otherwise prohibited from influencing the call's outcome. Target participants who spoke while their partners were using the phone would be sacrificing the success bonus for both participants. Callers were required to discuss any plans to use deception in the call ahead of time with their partners (the targets). Targets who felt uncomfortable with the progression of a call could terminate the experiment at any time and take over the conversation with their trustee—none did.

Whenever a caller reached a target's trustee, we asked the caller to complete a questionnaire about whether/how well they knew the trustee and how the call had proceeded. We considered the attack a success if the caller was able to get the trustee to either read the account-recovery code over the phone, forward the account-recovery code to the caller's email address, or forward the trustee-authentication email to the caller's email address. We considered the attack a failure if the trustee contacted the target to verify that the request was valid or failed to provide the caller with an account-recovery code or trustee-authentication email. Once the outcome of a call had been determined the trustee was allowed to speak with the participant, who in turn encouraged the trustee not to discuss the experiment with others.

Results

Despite our attempt to recruit pairs of participants who had varying levels of trust in one another, all but one of the 9 participant pairs were either spouses or had been friends for five or more years. Thus, we ended up with callers who were better acquainted with their targets and their targets' trustees than we had hoped to model in this experiment. However, we believe that the resulting data reflects the upper bound for attack efficacy.

Participants made calls that reached 49 total trustees, seven of whom were discarded because the trustee knew the target partner was at the study.

Of the remaining 42 calls, 19 (45%) resulted in the caller successfully obtaining either an account-recovery code or a trustee-authentication email. Of the remaining trustees, callers categorized 5 (12%) as being unwilling to provide the code, 14 (33%) as unable to provide it, and 4 (10%) as being too confused by the request.

Those callers in our experiment who already knew the trustee they were calling were more likely to succeed. The success rate for callers who knew the trustee they were calling was 14/25 (56%), compared to 5/17 (29%) for those who did not. While the result was not significant given the sample size, it certainly warrants further investigation.

In summary, telephone attacks by those well acquainted with the account holder posed a significant risk to our social authentication system. Once again, we hypothesize that informing trustees of their role ahead of time might help to ameliorate this problem.

We also hypothesize that this attack may become less effective if social authentication becomes commonplace, and users learn what to expect from a legitimate request.

DISCUSSION

This paper is by no means a comprehensive study of all attacks against our social-authentication system. While we studied two extremes of knowledge that attackers might have of the account holder, there is a large space in-between that we hope to study in the future.

There are many classes of attack that we have yet to study, including, but not limited to, the interception of trustee–authentication email, compromise of mail servers, denial of service, coercion, and the use of forged social-authentication sites to lure users to reveal information (phishing).

However, through our experiments we still discovered two important problems that must be addressed before our social-authentication system can be deployed:

1. Account holders cannot be expected to remember who they chose as trustees.
2. Current defenses are inadequate against phone-based attacks by close acquaintances.

To address both problems, we propose having trustees opt-into their role by responding to an email. These notification emails, and resulting conversations with the account holder, may provide opportunities to remind trustees that the account holder will never request account-recovery codes through third parties.

Notifying trustees of their selection by email could also have a negative impact on the system's security. Attackers who compromised an email account would not only gain the ability to act as a trustee with the victim's account, but could also search the victim's email archives to learn who had chosen the victim as a trustee.

As part of our ongoing effort to make social authentication practical, we will be testing the benefits and risks of opting trustees into their role in the near future.

CONCLUSION

We designed a last-resort social-authentication mechanism for websites, implemented the mechanism for Windows Live ID, and ran three experiments to test its reliability, efficiency, and security. We deployed the system within the domain space of a major web portal, `live.com`. We recorded the security behavior of trustees who were unaware that their actions could cause no harm, so as to increase ecological validity.

Our experiments identified two problems: most account holders forgot whom they had chosen as trustees and many trustees would reveal authentication-codes if called by someone close to the account holder. If we can address these problems, social authentication may be a valuable addition to websites' backup authentication toolbox: seventeen of nineteen participants who made the effort to reach at least one trustee were able to obtain the three trustee-authentication codes required to reset their password. Furthermore, the email attacks we simulated proved relatively ineffective at extracting codes from trustees.

ACKNOWLEDGEMENTS

Jeffrey Friedberg, from Microsoft's Trust User Experience team helped us navigate the privacy compliance process at Microsoft.

At Microsoft Research, A. J. Bernheim Brush helped with the design and planning our experiments. Trevor Eberl led the operation of our prototype and kept it running smoothly. Jon Howell, Jay Lorch, and the anonymous reviewers provided helpful feedback on earlier drafts.

REFERENCES

1. J. Brainard, A. Juels, R. L. Rivest, M. Szydlo, and M. Yung. Fourth-factor authentication: somebody you know. In *CCS '06: Proceedings of the 13th ACM Conference on Computer and Communications Security*, pages 168–178, New York, NY, USA, 2006. ACM.
2. T. Bridis. Hacker impersonated Palin, stole e-mail password, Sept. 18, 2008. Associated Press.
3. S. Brostoff and A. M. 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*, 2003.
4. CommonwealthBank. NetBank NetCode SMS, 2008. http://www.commbank.com.au/netbank/netcodesms/.
5. CREDANT Technologies. Mountains of mobiles left in the back of New York cabs, 16, 2008. http://www.credant.com/mountains-of-mobiles-left-in-the-back-of-new-york-cabs.html.
6. Google Inc. Contact Us - Google Accounts Help, 2008. http://www.google.com/support/accounts/bin/request.py?hl=en&contact_type=ara&ctx=accounts&uses_apps=no&product=other&submit=Continue.
7. M. Jakobsson, E. Stolterman, S. Wetzel, and L. Yang. Love and authentication. In *CHI '08: Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, pages 197–200, New York, NY, USA, 2008. ACM.
8. Microsoft Corporation. Complete the form below for Windows Live ID validation, 2008. https://support.live.com/eform.aspx?productKey=wlidvalidation&ct=eformcs&scrx=1.
9. J. Podd, J. Bunnell, and R. Henderson. Cost-effective computer security: Cognitive and associative passwords. In *OZCHI '96: Proceedings of the 6th Australian Conference on Computer-Human Interaction (OZCHI '96)*, page 304, Washington, DC, USA, 1996. IEEE Computer Society.
10. A. 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.
11. SafeNet, Inc. 2004 annual password survey results, 2005. http://www.safenet-inc.com/news/view.asp?news_ID=239.
12. S. Schechter, A. J. Bernheim Brush, and S. Egelman. It's no secret: Measuring the security and reliability of authentication via 'secret' questions. In submission.
13. K.-P. L. Vu, R. W. Proctor, A. Bhargav-Spantzel, B.-L. B. Tai, J. Cook, and E. E. Schultz. Improving password security and memorability to protect personal and organizational information. *Int. J. Hum.-Comput. Stud.*, 65(8):744–757, 2007.
14. M. Zviran and W. J. Haga. User authentication by cognitive passwords: an empirical assessment. In *JCIT: Proceedings of the Fifth Jerusalem Conference on Information technology*, pages 137–144, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.