RUST: A Retargetable Usability Testbed for Website Authentication Technologies

Maritza L. Johnson
Columbia University

Chaitanya Atreya*
Columbia University

Adam Aviv†
Columbia University

Mariana Raykova
Columbia University

Steven M. Bellovin
Columbia University

Gail Kaiser
Columbia University

Abstract

Website authentication technologies attempt to make the identity of a website clear to the user, by supplying information about the identity of the website. In practice however, usability issues can prevent users from correctly identifying the websites they are interacting with. To help identify usability issues we present RUST, a Retargetable USability Testbed for website authentication technologies. RUST is a testbed that consists of a test harness, which provides the ability to easily configure the environment for running usability study sessions, and a usability study design that evaluates usability based on spoofability, learnability, and acceptability. We present data collected by RUST and discuss preliminary results for two authentication technologies, Microsoft CardSpace and Verisign Secure Letterhead. Based on the data collected, we conclude that the testbed is useful for gathering data on a variety of technologies.

1 Introduction

The heightened interest in website authentication technologies is fueled by a rise in cybercrimes, such as phishing, and US federal regulations that require financial websites to use two-factor authentication [2, 7]. Website authentication technologies attempt to solve one direction of the mutual authentication problem on the Internet by either altering the login process or providing the user with supplemental information. The primary usability questions in website authentication are how does a website communicate to the user it is the real site and how does a user identify a malicious website? Usability plays a major factor in the effectiveness of the technology but receives little attention during development.

To facilitate usability evaluations, we present RUST, a Retargetable USability Testbed which is a testbed composed of a usability study design and a test harness for the test environment. First we discuss prior work in the area and then we describe the design process for the usability study. Next, we describe the test harness. Finally, we present results from two usability studies conducted at Columbia to illustrate how we validated RUST.

2 Background

Prior to website authentication tools gaining popularity, Whalen and Inkpen conducted a study to evaluate web browser security indicators [15]. They collected data on which indicators users considered when evaluating a webpage's security by asking participants to perform a set of tasks while focusing on the each website's security. Most participants checked for either the lock icon or https in the URL bar, but few checked for or understood certificates. Similarly, in an effort to understand why phishing attacks are successful, Dhamija et al. measured how users evaluate possible phishing websites [6]. Participants were presented with a series of websites and asked to determine if the site was real or fake. 23% of the participants based their decision solely on indicators they found in the webpage content. Some participants looked for the lock icon but mistook lock images in the webpage content for trusted security indicators.

Web Wallet is an anti-phishing tool that alters how password are entered [16] by providing an interface for entering sensitive information, other than the web form provided by the website. It helps the user by removing the guesswork of which websites have been visited in the past. A usability study showed Web Wallet was effective at helping the participants identify the real website but participants were easily tricked by spoofs of the interface. Wu et al. also evaluated the usability of toolbars to assess if they assisted users in identifying phishing websites [17]. The results of the usability study indicate the toolbars are ineffective in assisting users on well-designed spoofs. Another study evaluating website

*C. Atreya graduated from Columbia University in Fall 2007

†A. Aviv is currently a student at the University of Pennsylvania

authentication was Jackson et al.'s evaluation of whether browser indicators of extended validation certificates assisted users in identifying phishing attacks [11]. The results showed new indicators like a green URL bar for an EV certificate did not offer an advantage over the existing indicators.

Schechter et al. [13] conducted an in-lab study to evaluate a website authentication technology where each user has a personalized image. Participants were asked to perform a series of online banking tasks while security indicators were gradually removed. Their results show participants fail to recognize the absence of security indicators, like the SSL lock and HTTPS, and will enter their password in the absence of their personalized image.

Usability study design is a well-studied area [10], however, designing security usability studies creates additional challenges. One issue is how to design a study where the test administrators attack the participants [3]. More recently, usability studies have been designed to evaluate methods of conducting security usability studies. For example, Schechter et al. conducted a between-subject usability study to measure the effect of asking a participant to play a role and use fake credentials rather than their personal information [13]. They found participants who used their real data act more securely during the tasks. To help usability study designers, SOUPS made kits available from the papers in their proceedings [14]. The kits provide usability study material but are fairly specific and reusing the material would require a number of changes.

2.1 User Study Design

In RUST, usability is measured by the technology's spoofability, learnability, and acceptability. Spoofability is an attacker's ability to trick the participant into entering personal information on an illegitimate website. Learnability is the user's ability to correctly use the technology with and without instruction. Acceptability is the user's reaction to the technology; if users do not understand why a security process is necessary they will find ways to break the process [1].

We chose an in-lab study as the method of evaluation so we could attack our participants and measure spoofability without raising ethical concerns [12]. Because we wanted to see how participants would behave under conditions of attack, we did not disclose the purpose of the study beforehand, since doing so would place an unrealistic focus on security. We supplied participants with credentials to eliminate privacy concerns and because users do not already have the necessary credentials to use with the novel technologies we were testing. We asked participants to play a role during the session to justify the use of fake credentials and motivate the participant to act securely.

To evaluate both spoofability and usability, we asked participants to complete tasks at real and spoofed banking websites. Spoofability is measured by the number of successful attacks in a session. To evaluate learnability, four tasks are given before the participant is provided with instructions in the fifth task. This provides the chance to gather data on the technology's ease of use prior to the participant reading documentation. The acceptability of the technology is based purely on a participant's subjective opinion and cannot be measured through direct observation. Instead, we collected feedback through questions, using Likert scales to classify their reactions, and open-ended questions to comment on their thoughts during the session.

We designed the study as a within-subject study, where each participant is given the same set of tasks under the same conditions. We collected session data by the test harness and through self-reported feedback. Before beginning the study, we gave each participant a demographic survey with questions to gauge their experience with web browsing. We gave them copies of the study instructions, the role they are asked to play, and personal information to accompany the role. The instructions state the goal of the study is to improve online banking. We asked participants to imagine that they have an uncle who is in the hospital for an unexpected extended period of time and needs someone to assist in managing his finances. In addition, we asked the participant to act normally and to treat their uncle's information as they would their own.

During a session, we sent the participant eight emails, each of which contains a task and a link to the website where they should complete the task. Four of the emails are phishing attacks and direct the participant to an illegitimate site. The other four emails direct the participant to a real financial institution's website. Some of them are requests from Uncle John for a specific action to be taken, and one is an email from the bank introducing the new technology and providing basic instructions. The first task directs the participant to the real site, and allows them to experience the technology working properly before an attempted spoofing attack. Between each of the tasks, we asked the participants to comment on their experience if they completed the task, or why they decided not to complete the task for any reason. After the tasks are completed we asked participants to express their opinion of the technology in a post-study questionnaire.

2.2 Test Harness Implementation

The test harness component of the RUST testbed creates a transparent testing environment that can be easily

configured for different technologies, thus allowing for a simplified process for conducting usability evaluations. The use of proxies and logging tools allow for data to be collected on the participant's actions while providing the environment for serving the necessary study webpages. RUST collects data by monitoring the URLs requested while keeping track of the order that pages are accessed, which indicates whether a phishing attack was successful. Additionally, the test harness monitors the time spent on a webpage. Time monitoring is important for a technology with multiple steps or pages because the time spent on a page may indicate possible problems in the login sequence. To monitor the order the webpages are visited and the time spent on each page, a JavaScript beacon is inserted on each test webpage the participant may visit. The test harness also includes scripts to convert the logging output into a more manageable format. A Python based script is used to send MIME based emails during the test session. Also, a simple shell script is included in the testbed to send specific emails to participants at fixed time intervals during the session. The complete details of the design are contained in [4].

3 Evaluating RUST

We used RUST to conduct usability study sessions with Windows CardSpace and Verisign Secure Letterhead. Each of the technologies approach the website authentication problem differently. CardSpace changes the login procedure, while Secure Letterhead provides the user with additional information to help them determine the website's identity.

For recruitment we solicited participants on Craigslist, placed a listing in the Facebook Marketplace and posted fliers at Columbia University. The only requirement that we places on participants is that they had experience with web browsing.

3.1 Windows CardSpace

CardSpace is an identity metasystem that manages a user's online identities [5], replacing usernames and passwords. When a user needs to log in to a website, the user clicks on a button in the page content that begins the login process. At this point, CardSpace is launched and the website's credentials are sent. When the CardSpace interface appears, the user selects the appropriate "card" that represents their credentials. The user does not submit any personal information to the website. Instead, the exchange of credentials is handled completely within the CardSpace protocol once the "card" representing the user's identity is chosen by the user. Since the protocol is designed to reveal information only to the verified parties, tricking the user into giving away their credentials is less feasible and different attacks are required. As a result, we designed spoofs that focus on the user enterinig sensitive information directly to the website. One email stated CardSpace was down for scheduled maintenance and directed the user to a spoofed page to signup for temporary access. The user was then sent to a webpage requesting sensitive information. The other spoof redirected the user to a form when they attempted to login with CardSpace, stating the participant must register for an identity card by entering personal information. Since the user logged into the same institution's website with a identity card previously, this should be alarming.

3.1.1 Data Collected

We recruited a total of 13 participants to evaluate CardSpace, 4 female and 9 male. Their ages ranged from 18-60 and each of them spend 20 or more hours per week on the Internet.

Of the eight tasks given in a session, the four tasks prior to the CardSpace instructional email directed the participant to: a real website with CardSpace, a spoofed site that asked the participant to register for a new identity card, a real website with CardSpace, and a spoofed website that stated CardSpace was down for maintenance. The four remaining tasks were in the same order and the first of the second set of four was also the instructional email.

In the first task, on the real website, 12 of the 13 participants reported some level of confusion. However, 6 participants stated in the post-study questionnaire that CardSpace was intuitive, despite commenting they were confused after the first task. Their comments after the first task include: *"It took me a little while to know where I was suppose to click", "I wasn't asked to enter account info which was suspicious", "I understood the task, but the first thing I should have been asked is my account id and password", "when I clicked on the log-in tab I was a little confused, I am accustomed to seeing a user ID box and then some type of password, but what popped up were cards, there seems to be some serious security lapse", and "when I clicked login something irrelevant came up but when I hit ok I was logged in".*

In the first spoof, when participants were told to register for a new card, 11 participants completed the task without noticing anything suspicious. One person realized they had been tricked when they were redirected to the real site but this was after entering personal information. In the fourth task when the spoofed website stated CardSpace was "down for maintenance", the same 11 participants fell for the spoof. The participant who realized they were spoofed previously recognized the spoof without entering any information. After reading the in-

structional email, two people reported they were still confused when completing the task that followed on the real site. By the fourth task, the second task on a real website, no one commented they were confused.

After reading the instructional email, 3 of the 13 participants did not fall for the spoof that occurred immediately after (a request to register for a new card). One participant was previously an identity theft victim and the other two participants cited the instructional email as their reason for not completing the task. The final spoof stated CardSpace was down for maintenance, and 3 people were not tricked by it. The identity theft victim was not tricked and one person cited the instructions. However, one person who cited the instructions in the previous spoof fell for the second spoof. The other person remarked the site didn't look right in general, so they refused to complete the task. In the post-study questionnaire, four people said the CardSpace login procedure required slightly too much time and six reported the amount of time required by CardSpace was just right.

Overall, the participants were mostly confused that CardSpace took the place of a username and password. The spoof that stated CardSpace was down for maintenance tricked 11 of the 13 participants and after the instructional email the same spoof tricked 10 of the 13. The spoof that asked the participant to register for another card was successful on 11 of the 13 participants and after the instructional email successfully tricked 10 of the 13.

3.2 Verisign Secure Letterhead

Secure letterhead assists the user by displaying security context information interactively and more prominently in the web browser's primary interface. The implementation we tested was Firefox extension and displayed the logotype and certificate authority fields from an extended validation X.509 certificate [8, 9]. When the user navigates to a website that has an extended validation certificate, the logotype is displayed in the upper right hand corner of the browser next to the location bar. It extracts relevant fields from the SSL certificate, which can be difficult for users to find and understand. The fields are displayed as an interactive visual indicator. More certificate information is shown when user clicks on the logotype.

3.2.1 Data Collected

We conducted five sessions with Secure Letterhead, two female and three male. Their ages ranged from 18-50 and all reported to spend 20 or more hours on the Internet per week. Only one of the users demonstrated prior knowledge of phishing.

The four tasks prior to the instructional email that detailed how to use Secure Letterhead directed the user to a real site, a fake site where no additional information was displayed, a real site, and a fake site spoofing Secure Letterhead with CSS and HTML. The four remaining task emails were in the same order except the first of the second set was also the instructional email.

Since Secure Letterhead does not alter the login process, the first task went smoothly for the participants, however, no one noticed the presence of the logotype or tried to click on it. The first spoof, where no additional information was presented, tricked all 5 participants. One participant realized they had been tricked and wasn't tricked again during the study. However, Secure Letterhead did not play a role in their realization; instead they realized they were tricked after logging in and being forwarded to the real login page. Again, the task on the real website went smoothly, but no one noticed the logotype. The second spoof, where the interface was recreated using CSS and HTML, was successful in tricking 4 of the 5 participants.

Every participant attempted to access the information in the logotype after reading the instructional email. One commented the instructions were incorrect and the logo was on the left, indicating they clearly misunderstood which logo the instructions referred to and mistook an image in the webpage content for the logotype. Another did not to complete the task and wrote they didn't have a reason to trust the information it displayed. Another commented the instructions were too wordy, then during the task attempted to access the information by clicking on the logo in the page content. This behavior suggests users are unable to distinguish between page content and trusted indicators, which was also observed by Whalen et al. [15] and Dhamija et al. [6].

The first spoof after the instructional email, where no additional information was displayed, tricked 4 of the 5 participants. In the remaining three tasks, only two attempted to access the information. One of them attempted to access the information after being redirected to the real site because they had already entered their credentials on the illegitimate site. In other words, they checked the credentials at the wrong point of interaction and had already lost their password. The other one checked the logotype on the real website. The final spoof with the recreation of Secure Letterhead in content successfully tricked 4 of the 5 participants.

In the post-study questionnaire, 3 of the 5 participants reported they would remember to check for the logotype if they were about to do something important. 2 of the 5 participants stated they would not have figured out how to use Secure Letterhead without instructions. One person commented the information displayed did not seem useful, and two commented the information might be

useful but may not be necessary.

Overall, it appeared as though the participants were unsure of how to interpret the information presented by Secure Letterhead, why it was important, or why they should trust it. These comments were made in the post-task questionnaire after receiving the instructional email and in the post-study questionnaire.

4 Conclusions

This paper presents RUST, a testbed for evaluating the usability of website authentication technologies. The results we present demonstrate the versatility of RUST, its ability to test different types of technologies, and the detailed feedback it collects about why participants are tricked, which would not be possible in an in-the-wild study. Though RUST is not intended to measure how users' performance is affected by time, minor changes can be made to account for this new goal. As an additional benefit, RUST can be used to compare technologies since they same usability study design can easily be used to evaluate different technologies.

5 Acknowledgments

This work was supported by the Financial Services Technology Consortium. The authors would like to thank the AFIC project participants for their feedback and the UPSEC reviewers for their comments.

References

[1] ADAMS, A., AND SASSE, M. A. Users are not the enemy. *Commun. ACM 42*, 12 (1999), 40–46.

[2] APWG. Anti-phishing working group. `http://www.antiphishing.org`.

[3] ARSHAD, F., AND REEDER, R. When user studies attack: Evaluating security by intentionally attacking users. SOUPS Conference Report `http://www.ieee-security.org/Cipher/ConfReports/2005/CR2005-SOUPS.html#_Toc110303259`.

[4] ATREYA, C., AVIV, A., JOHNSON, M., RAYKOVA, M., BELLOVIN, S. M., AND KAISER, G. Rust: The reusable security toolkit. In *submission to SOUPS '08: Proceedings of the symposium on Usable privacy and security* (2008).

[5] BROWN, K. Step by step guide to infocard. `http://msdn.microsoft.com/msdnmag/issues/06/05/securitybriefs/default.aspx`.

[6] DHAMIJA, R., TYGAR, J. D., AND HEARST, M. Why phishing works. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems* (NY, NY, USA, 2006).

[7] FFIEC. Authentication in an internet banking environment. `http://www.ffiec.gov/pdf/authentication_guidance.pdf`, 2005.

[8] FORUM, C. A. Extended validation SSL certificates. `http://cabforum.org/`.

[9] HALLAM-BAKER, P. Secure internet letterhead. In *W3C Workshop on Transparency and Usability of Web Authentication* (2006).

[10] ISO/IEC. ISO 9241-11 guidance on usability.

[11] JACKSON, C., SIMON, D. R., TAN, D. S., AND BARTH, A. An evaluation of extended validation and picture-in-picture attacks. In *Proceedings of Usable Security (USEC '07) Workshop* (2007).

[12] JAKOBSSON, M., AND RATKIEWICZ, J. Designing ethical phishing experiments: A study of (ROT13) rOnl query features. In *WWW '06: Proceedings of the 15th international conference on World Wide Web* (NY, NY, USA, 2006), ACM.

[13] SCHECHTER, S. E., DHAMIJA, R., OZMENT, A., AND FISCHER, I. The emperor's new security indicators. In *Proceedings of the 2007 IEEE Symposium on Security and Privacy* (Washington, DC, USA, 2007), IEEE Computer Society.

[14] SOUPS 2006. Security user study toolkits. `http://cups.cs.cmu.edu/soups/2006/workshop-kits/kits.html`.

[15] WHALEN, T., AND INKPEN, K. M. Gathering evidence: Use of visual security cues in web browsers. In *GI '05: Proceedings of the 2005 conference on Graphics interface* (Waterloo, Ontario, Canada, 2005), Canadian Human-Computer Communications Society.

[16] WU, M., MILLER, R. C., AND GARFINKEL, S. L. Do security toolbars actually prevent phishing attacks? In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems* (NY, NY, USA, 2006), ACM Press.

[17] WU, M., MILLER, R. C., AND LITTLE, G. Web wallet: Preventing phishing attacks by revealing user intentions. In *SOUPS '06: Proceedings of the second symposium on Usable privacy and security* (New York, NY, USA), ACM Press.

A Windows Cardspace

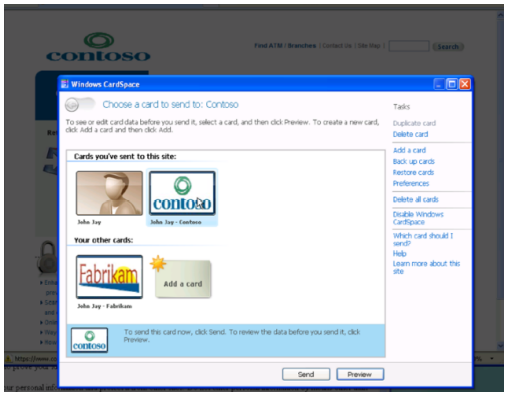

Figure 1: The user interface for selecting an identity card.

B Verisign Secure Letterhead

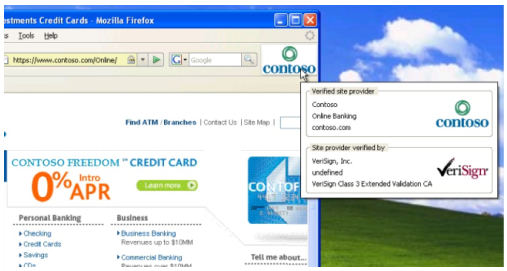

Figure 2: The Extended Validation certificate fields displayed by the Firefox extension when the logotype is moused over.

C Verisign Secure Letterhead Spoof

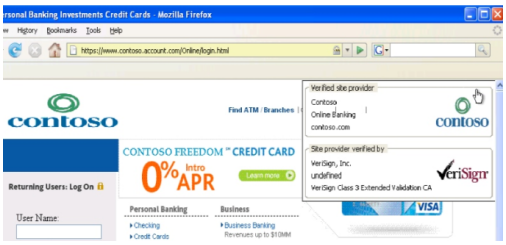

Figure 3: Verisign Secure Letterhead recreated in the webpage content using CSS and HTML.

D User Study Material

This section provides the questionnaires from RUST's usability study.

D.1 Post-task Questionnaire

The participant is given a post-task questionnaire for each task in the session. Each questionnaire has a set of questions to answer if the participant completed the task and a set of questions to answer if they chose not to complete the task.

D.1.1 Questions for those who completed the task

If you decided to complete the task, please answer the following questions:

1. Please mark the most accurate description of your experience with this task:
   - I knew what to do right away.
   - After a moment I knew what to do.
   - After a few minutes I knew what to do.
   - I completed the task, but I'm not sure I did it correctly.
   - Something was unclear about the task and I was unable to figure out what to do.
2. Additional comments: (eg: about the task, email, website, an explanation of your previous answer)

D.1.2 Questions for those who did not complete the task

If you decided *NOT* to complete the task, please answer the following questions:

1. What contributed to your decision to not complete the task?
2. Additional comments: (eg: about the task, email, website, an explanation of your previous answer)

D.2 Post-study Questionnaire

1. When a site asks you for sensitive personal information (eg: account information, passwords, social security number), which of the following do you check to ensure you're giving your information to a valid website? (When you think you're on your bank's website, how do you know that you aren't on another site?)

   Check all the apply:

- Look for the lock icon
- Check the url in the address bar
- Check the certificate of the page
- Look at the information on the page (the images, the text, or the brand name)
- Look for the security or privacy policy
- Wait for personalized image or text to be displayed (not available on all websites)
- Other, please specify.

2. What do you typically do when you receive an email that provides you with a link and asks you to go to the website and sign into your account, to either update your account or verify a transaction?

   Check all the apply:

   - Click on the link to follow the instructions
   - Type the url into the address bar to follow the instructions
   - Go to the site using a bookmark to follow the instructions
   - Delete the email
   - I don't receive emails like this
   - Other, please specify

The goal of the technology evaluated in this study is to give the user a better way of knowing they are on the correct site. During the study you should have also received an email with instructions on how to correctly use Name of technology.

Technology : Brief description of technology and how it should work, similar to the instructional email.

Please refer back to the above description as needed for the remaining questions. If anything is unclear, please ask questions.

1. Prior to reading the instructional email:

   - Was it noticeable that the technology was in place? If so, what was noticeable?
   - Did the technology make it clear how it should be used? If so, please describe.

2. Imagine you were required to use *technology* each time you signed in to a financial institution's website. Choose one of the following to describe your feeling about the amount of time it requires:

   - Not enough
   - Slightly too little
   - The perfect amount
   - Slightly too much, but reasonable
   - Way too much
   - Comments:

3. Choose one of the following to describe how often you would remember to follow the process closely:

   - Never
   - When I was about to do something important
   - Most of the time
   - Every time
   - Comments:

4. Choose one of the following to describe your overall impression of *technology*:

   - The information it provides is useful, and it seems necessary.
   - The information it provides appears useful, but I'm not sure it's necessary
   - The information it provides does not appear useful.
   - The information it provides is not useful to me at all.
   - Other, please explain.
   - Comments