The Johnny 2 Standardized Secure Messaging Scenario

Simson L. Garfinkel
MIT CSAIL
Cambridge, MA 02139
simsong@csail.mit.edu

Robert C. Miller
MIT CSAIL
Cambridge, MA 02139
rcm@csail.mit.edu

ABSTRACT

We present a scenario for user testing secure messaging tools and anti-phishing technology. The scenario, *Johnny 2*, is loosely based on the scenario that Whitten and Tygar presented in their acclaimed paper "Why Johnny Can't Encrypt,"[14], but provided with significantly refined detail and automation. We recently used this scenario successfully in a user test with 43-subjects. We hope that by developing and using standardized user test scenarios, research conducted by researchers in the field of usability and security can be more easily compared and contrasted.

Categories and Subject Descriptors

D.4.6.c [**Security and Privacy Protection**]: Cryptographic Controls; H.5.2.e [**HCI User Interfaces**]: Evaluation/methodology

General Terms

Usability, Security

Keywords

User Studies, E-Commerce, User Interaction Design

1. INTRODUCTION

One of the most effective ways to test the usability of a security tool is to put a human subject into a scenario where the tool must be used to fulfill a simulated job function. But while these studies can be very rewarding for the experimenter, they can also be very difficult to construct. The scenario approach relies on the willingness of the subjects to "suspend their disbelief" and play the role that they have been assigned. Inconsistent details and implausible explanations can be jarring to the subject, interrupt the scenario, and poison the results.

As the number of researchers in the field of Human Computer Interaction Security (HCI-SEC) increases, it would be useful to have a library of high-fidelity scenarios. These scenarios could then be used to test new security tools, making it easier to compare the results of different tools that were subject to human testing. Such a library of scenarios could also be pre-approved by Internal Review Boards at colleges and universities, making it possible for students to perform actual human subject testing in a typical introductory course without needing to file any paperwork.

SOUPS '05 Pittsburgh, Penn. USA

1.1 The Johnny Scenario

In 1999 Carnegie Mellon University graduate student Alma Whitten and her advisor J. D. Tygar published "Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0."[14] That paper reports on a user study in which Whitten asked 12 subjects to create keys and send digitally signed and sealed messages using the Macintosh version of PGP 5.0 and Eudora. To make the task more realistic, Whitten and Tygar invented a scenario in which the human subjects played the role of a volunteer in a political campaign. Their mission was to use the PGP program to create a PGP public key, have the key signed by another member of the campaign, and then use the key to send the other campaign member e-mail that was both digitally signed and sealed. Most of the subjects failed in this task.

Since 1999, *Johnny* has become one of those most oft-cited papers in the field of usability and security (HCI-SEC). Most citations concentrate on the paper's primary findings, that programs like PGP 5.0 can have attractive user interfaces but nevertheless be difficult to use from a security prospective.

As part of a larger project involving usability and security, we decided to replicate the *Johnny* protocol but to substitute Microsoft's Outlook express with S/MIME for Eudora with PGP 5.0. Our goal was to see if we could overcome the usability failings that Whitten and Tygar had reported by using modern software with a different key certification model.

However, in plumbing the depths of the author's 15-page Usenix paper, their 39-page CMU technical report[13], and Whitten's 229-page dissertation[11], we discovered that many details of the experiment that we thought were necessary to replicate the Johnny protocol were missing. Moreover, aspects of the original *Johnny* protocol did not lend themselves to automation.

After contacting Whitten and learning that key aspects of the 1998 study had not been preserved, we created a new experimental protocol, *Johnny 2*, for the purpose of testing the Key Continuity Management model suggested by Gutman.[5]

This paper presents the *Johnny 2* protocol in detail, discussing its strengths, weaknesses, and areas of possible improvement. It does not present the results of our usability tests, which have been submitted elsewhere.[4]

1.2 The Need for Standardized Scenarios

We feel that the *Johnny 2* protocol is an excellent candidate to become a standardized HCI-SEC scenario for researchers working in the area of secure messaging, spoofing, phishing, and even basic PKI work.

We believe that the use of standardized HCI-SEC scenarios will have several key benefits, including:

- Researchers will able to devote less time to developing scenarios and more time developing the specific technology that they wish to test.
- Researchers can spend more time presenting their findings and less discussing their scenarios.
- It will be easier to compare results of different studies.

While the *Johnny 2* scenario is certainly not sufficient for testing *all* end-user security tools, we believe that it can be adapted to testing many such tools. We hope that researchers working in this area will consider using this scenario and that researchers in other HCI-SEC fields will join us in developing a library of scenarios that can be used by other researchers throughout the community.

1.3 Related Work

We are unaware of any standardized scenarios or protocols that are used in either the HCI-SEC community or the greater HCI or security communities.

Other disciplines, however, make significant use of standardized scenarios or experimental protocols for training, and subject evaluation. Medical schools use standardized patient scenarios for training student physicians (*e.g.* [3]) and testing practicing physicians (*e.g.* [1]). Social psychologists have long used standardized tests (*e.g.* [8]) for categorizing participants in research studies.

2. THE JOHNNY 2 SCENARIO

The scenario in the original *Johnny* paper was interesting and straightforward: the experimental participant has shown up for work the first day as a volunteer at a political campaign that is trying to get a candidate elected to some state-wide office in Pennsylvania. The volunteer has been assigned the role of Campaign Coordinator and is responsible for sending the candidate's schedule to members of the campaign team. Quoting from Whitten's "Initial Briefing" document:

> "It is very important that the plan updates be kept secret from everyone other than the members of the campaign team, and also that the team members can be sure that the updates they receive haven't been forged. In order to ensure this, you and the other team members will need to use PGP to encrypt and digitally sign your email messages."[13, p.38]

Johnny 2 clarifies and further develops both the Campaign and the Attacker. To make the task seem more realistic, the individual campaign members are given roles and backstories.(See Table 1.) The Attacker is now a clearly defined member of the opposing campaign with specific capabilities and means of action. Communication channels are clearly proscribed. Finally, the *Johnny 2* scenario is designed to allow easy automation, making it possible for a single experimenter to easily run a subject in an hour. It is even possible for the scenario to be run in a completely autonomous manner—for example, completely over the web using a subject's computer at a place and time of the subject's choosing.

Automation is accomplished through the use of eight pre-scripted messages that are sent to the subject by the experimenter. The messages are drafted in such a way that they can be sent simply by clicking a button: the experimental subject's answers do not determine the next message to be sent and do not have an impact on the message contents. The *Johnny 2* scenario features internal controls that can be used to gauge the competence of the experimental subject in an objective manner. Finally, the *Johnny 2* scenario introduces a series of escalating attacks that model the kinds of attacks that are common on the Internet today. These attacks could easily be substituted with other attacks by experimenters interested in exploring other HCI-SEC issues.

2.1 The Backstory

In the *Johnny 2* scenario, the fictional campaign team has decided to equip its computers with a new security tool that we shall call $\mathcal{ST}$. In our work, $\mathcal{ST}$ was a system called CoPilot that implemented Key Continuity Management, but $\mathcal{ST}$ could be any system that provides some form of secure messaging functionality. Alternatively, $\mathcal{ST}$ could be program such as SpoofGuard[2] that is designed to provide anti-phishing protection for a web-based messaging system such as HotMail.

Ben Donnelly, the Campaign's IT coordinator, has loaded $\mathcal{ST}$ onto the the Campaign's office computer, which is the same computer that the subject will use for the user test experiment. (If loading $\mathcal{ST}$ is part of the user test, then Ben Donnelly has provided the subject with a copy of the program on a CDROM.)

The security tool is designed to facilitate the transmission of a "secret" to some members of the Campaign while preventing the dissemination of the secret to the public or members of the opposing campaign. This is the task on which the subject is tested.

2.2 The Initial Task Description

The *Johnny 2* protocol is designed so that subjects can begin the experiment with minimal training. In our case, training consisted of a single paragraph of text inserted into a human subject consent form and a few sentences provided to the subject on a page document entitled "Initial Task Description."

The language of the single paragraph insert into the consent form was taken from Whitten's consent form and then modified to substitute the word "CoPilot" for the original "PGP:"

> In the test, you will be asked to play the role of a volunteer in a political campaign. After you volunteered, you were given the role of Campaign Coordinator. Your task is to send updates about the campaign plan out to the members of the campaign team by email. It is very important that the plan updates be kept secret from everyone other than the members of the campaign team, and also that the team members can be sure that the updates they receive havent been forged. In order to ensure this, you and the other team members will need to use CoPilot to make sure that all of the email messages are secure.

The relevant portions of the "Initial Task Description" document appear in Figure 1; these were taken from Whitten's original "Initial Task Description," as reprinted in [13].

Placing important training information inside the human subject consent form may be considered by some people to be a mistake. Many of the people participated in our user study appeared to be serial human subjects who participate in many studies on the MIT campus and have apparently been conditioned to ignore the verbiage contained in federally-mandated consent forms. We attempted to avoid this problem by reading every word of the consent form to the subjects. On several occasions we had to tell the subjects words to the effect of "this isn't a standard consent form, you need to read it in order to participate."

Experimental Subject:		
Campaign Coordinator	ccord@campaign.ex.com	Experimental subjects are told: "You are the Campaign Coordinator."
Campaign Personnel:		
Maria Page	mpage@campaign.ex.com	Campaign Manager and the Coordinator's boss.
Paul Butler	butler@campaign.ex.com	Campaign finance manager.
Ben Donnelly	bend@campaign.ex.com	IT coordinator. Officially Paul's assistant, but also a full-time student at the University of Pennsylvania.
Sarah Carson	carson@campaign.ex.com	"A full-time graphics designer."
Dana McIntyre	dmi@campaign.ex.com	Office manager, but away for the week because her husband is having surgery. (Don't worry, it's a routine procedure.)
Attacker:		
Attacker Paul	butler@campaign.ex.com	Claims to be Paul Butler, having computer problems.
Attacker Sarah	sara_carson_personal@hotmail.com	Claims to be Sarah Carson, sending email from home using her "personal Hotmail account" because she can't get to her campaign email from home.
Attacker Maria	mpage@campaign.ex.com	Attacker "Maria" sends an unsigned message to the Campaign Coordinator asking that the schedule be sent to both Ben and Sarah.

Table 1: Personas used in the *Johnny 2* experiment.

Initial Task Description

You are the campaign coordinator.

You are working for the campaign manager, Maria Page, mpage@campaign.ex.com

The other members of the campaign team are:

Paul Butler, `butler@campaign.ex.com`
Ben Donnelly, `bend@campaign.ex.com`
Sarah Carson, `carson@campaign.ex.com`
Dana McIntyre, `dmi@campaign.ex.com`

You have arrived early for work. No one else from the campaign is in the office.

If you wish to use the telephone to call a campaign member, please ask the experimenter for a "phone."

When you are asked by Maria, please send the schedule to the other team members.

Once you have done this, wait for any email responses from the team members, and follow any directions they give you.

Don't forget to "think aloud" as much as you can.

Figure 1: The Johnny 2 Initial Task Description

On the other hand, placement of this text inside the consent form has the advantages that the subjects were given something else to do and think about after they were told that security was critical in the experiment. This may ultimately result in more realistic reactions and performance on the part of the test subjects.

2.3 The Johnny 2 Attacker

The Attacker is an affiliate with the opposing campaign who is determined to use trickery to obtain the schedule, but is not willing to violate any laws—that could easily backfire. Thus, the Attacker is not going to try to break into the candidate's email server or call up the candidate's ISP and attempt to have the password on the candidate's email account reset. Instead, the Attacker merely tries to trick the Campaign Coordinator into revealing the candidate's secret campaign schedule.

Equipped with knowledge of the campaign's personnel (perhaps obtained by previously calling up the campaign and getting a list of the workers), the Attacker creates a series of Hotmail accounts with names that are similar to members of the campaign. The attacker then sends email to the Campaign Coordinator claiming to be a member of the Campaign, complaining that there is a problem with the Campaign's email system, and asking the Coordinator to email the coveted schedule to one or more of the Hotmail accounts. (Mitnick outlines this style of attack [7], and asserts that it is commonly used in the field of social engineering attacks.)

These messages constitute an escalating attack that can be used to gauge the effectiveness of various defenses offered by *ST*. The attacks could be mounted sequentially, in randomized order, or different subjects could receive different attacks.

As it just so happens, the Attacker's messages are sent on the Campaign Coordinator's first day on the job. And what is the Campaign Coordinator doing this first day? The Coordinator is sending out copies of the Candidate's coveted schedule to all of the members of the campaign team, each message sent at the request of Campaign Manager Maria Page. Due to a variety of circumstances, no other member of the Campaign team is in the office.

Not content with simple trickery, the Attacker attempts to maximize his chances of success by jamming the Campaign's telephone lines. Such attacks are actually quite easy to do, and have in fact been carried out in the past during actual political campaigns—for example, such an attack was carried out against New Hampshire's state Democratic party in an attempt to counteract the party's get-out-the-vote effort on Election Day 2002.[9]

2.4 The Johnny 2 Messages

During the experiment the experimenter sends eight messages to the experimental subject. Each message claims to be from a Campaign personna. Message #1 is from Maria Page, the Campaign Manager and the experimental subject's fictional boss. This message orients the subject and gives the subject information about the other Campaign members. Message #2 gives the subject an instruction to send confidential information to two of the campaign members. Message #3 is a request from a campaign member for

a copy of the confidential information. According to the briefing that the experimental subject has received, requests for confidential information made by legitimate campaign members should be satisfied.

In Message #4, the Attacker claims to be Paul Butler, the campaign's finance manager. We say that this message is sent from Attacker Paul to distinguish the message from a message that might actually be sent from the "real" Paul Butler. Message #4 appears to be sent from Butler's campaign e-mail address, but in fact this address has been spoofed. Butler says that he is having problems with his computer and asks that the secret be sent to both his Campaign and personal email addresses.

Message #5 claims to be from Sarah Carson, the Campaign's graphic designer. Attacker Sarah says that she is working from home and does not have access to her Campaign email. (The subject hasn't been told if campaign workers can or cannot access their campaign email from home.) She asks that the secret be sent to her personal email address.

Message #6 claims to be from Maria Page, but in fact is sent by the attacker. In this message Attacker Maria says that she has spoken by phone with Paul and Sarah, they need the secret, and the experimental subject *must send the message now!* Right now! Do it! This is your boss speaking!

The final two messages are from the "real" Maria Page. In message #7 Maria asks that the secret be sent to two other campaign members, while message #8 announces that the test is over.

These messages are summarized in Table 2.

3. EXPERIMENTAL APPARATUS

Although there is a temptation to perform user testing in an expensive facility with one-way mirrors or hidden video cameras, we found that we were able to conduct a very successful user test in a typical graduate student office. The test subject was given use of a standard Dell computer with a 17-inch LCD screen while the experimenter controlled the experiment and took notes on a Macintosh Powerbook (Figure 2). Care was taken so that the experimental subject could not see the contents of the laptop's screen (Figure 3).

3.1 Camtasia Studio

Whereas Whitten used a camcorder to videotape her sessions, we found that the screen recorder program Camtasia Studio offered by TechSmith [10] proved a better solution. Camtasia is a windows-based application that simultaneously records the contents of the computer's screen an whatever audio is presented at the computer's microphone. Camtasia then combines the video and the audio to produce a full-motion video that can be edited or played back on a PC running Windows or a Macintosh running MacOS 10. Camtasia is available as a free 30-day download, making it accessible for student use, and low-cost academic pricing is available.

There are many advantages to Camtasia over a traditional video recorder. The primary advantage is that the digital video recording is significantly easier to work with than the video tapes that a camcorder would produce. They are nevertheless compact: our 43 user trials required less than 13 gigabytes of storage. The screen is recorded with a lossless compression system, making it easy to capture screen shots for use in publications. Finally, recording the screen and the audio but not the user's face, does an excellent job at helping to preserve the subject's privacy — a requirement for research involving humans at federally-funded institutions.

There are two problems that we encountered using Camtasia. The first is that the screen recording system used by Camtasia caused the computer's creen to freeze for a few tens of a second every second; according to the documentation such freezing is the result of the screen capture process and may be less noticeable with other video cards. The second problem that we encountered is that the video Camtasia creates is compressed with a special codec created by TechSmith; in order to play this video back on another computer it is necessary to first download and install the codec.

TechSmith sells a more expensive screen recording system called Morae that is specifically designed for performing user interaction studies. The system records every mouse click and keystroke, and has provisions for an optional video camera trained at the user for capturing facial expressions. This product looks promising, although the video recording, by its very nature, is more invasive.

3.2 The Campaign Phone

Early in our pre-test several subjects commented that if the scenario had been real, they would have attempted to use a telephone to call one of the campaign members. At this point we realized that we needed to provide some kind of "phone" for the experimental subjects. The phone, shown in Figure 4, gives the subjects an opportunity to ask for a phone without interrupting the flow of the experiment. We found it useful to track if the subject had asked for a phone by entering the word "PHONE" into the Experimenter's Work Bench, described in the next section.

3.3 The Johnny 2 Experimenter's Work Bench

It is apparent from reading Whitten's reports and thesis that messages sent to test subjects during the *Johnny* trial were composed interactively during the experiment and sent by the experimenter.[1] This approach was rejected out-of-hand for *Johnny 2* for several reasons, including:

- Composing messages during the trial could lead to mistakes such as typographical errors, messages being sent to the wrong address, messages being sent without encryption or signing, and so on.
- If different test subjects received different messages, it would be difficult to perform anything but a qualtative analysis on the research findings.
- Given the need for the experimenter to be taking notes, the added overhead of writing detailed replies to email messages would have been very demanding.
- If the experimenter was obviously responding to the subject's email, the experiment would have lost considerably verisimilitude.

Instead, a program called the "*Johnny 2* Experimenter's Work Bench" was created for for administrating the experiment (see Figure 5). This program consisted of a graphical user interface running on the experimenter's Macintosh computer and two underlying programs, `sendmessage` and `send_signed`, that performed the

[1]Although the Whitten's writings contain many technical details of the *Johnny* experiment, notably missing are the actual messages that were sent by the experimenter to the study participants. In December 2004 Whitten was contacted and asked for a copy of the messages. Whitten responded that she had not retained them, but recalled that the messages were "pretty minimal" and consisted of little more than a three-message sequence:

1. "I couldn't decrypt that message, something must be wrong."
2. "I still can't decrypt that message, something is still wrong."
3. "I still can't decrypt that message, are you using my key to encrypt?"[12]

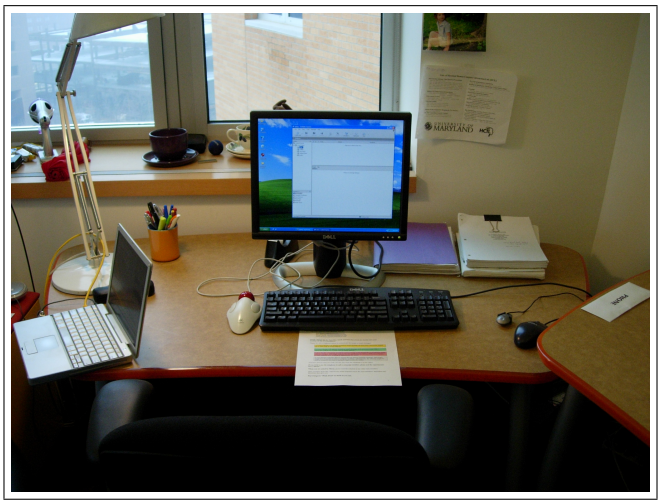

Figure 2: A photograph of the Johnny 2 experimental station. The experimenter's laptop is visible on the left. In front of the keyboard is the Johnny 2 "Initial Task Description." At the right is the "PHONE" (see Figure 4.)

Figure 3: A view of the experimenter's laptop from the experimental subject's chair. Note that the laptop's screen is not visible.

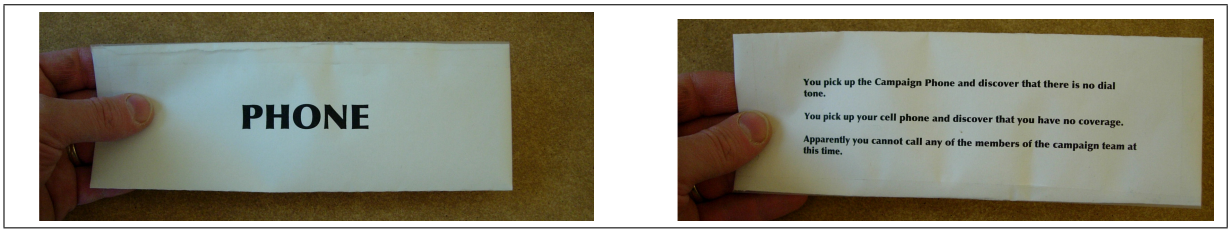

Figure 4: The front and back of the Johnny 2 "PHONE". The text reads: "You pick up the Campaign Phone and discover that there is no dial tone. / You pick up your cell phone and discover that you have no coverage. / Apparently you cannot call any of the members of the campaign team at this time."

actual business of sending email messages to the subject. The work bench program gives the experimenter a place to make notes, automatically timestamping them each minute and noting when each test message is sent to the subject.

Prior to the start of the experiment, the subject was told that the experimenter would be typing whatever they said, taking notes, and "sitting here reading my email" during the course of the experiment.

In practice, the experimenter was able to create a running transcript of the experimental subject's statements as they were made, and there was little need to refer to the Camtasia recording after the trail. What's more, the experimenter's constant typing made it easier for the experimental subject to ignore the experimenter, as there was no obvious correlation between what the subject was doing and what the experimenter was doing.

The automation provided by the Experimenter's Work Bench combined with Camtasia made it possible for a single experimenter to run as many as six subjects in a typical 8-hour workday. This makes it possible to run many more subjects than would typically be the case with a human subject study, which increases the chances of obtaining data that is statistically-significant.

4. RECRUITMENT

Whitten and Tygar recruited their subjects with posters that were put up at CMU and email messages that were sent to a variety of CMU electronic bulletin boards. A typical recruitment poster, taken from Whitten's dissertation, appears in Figure 6,

Because we originally intended to use *Johnny* as a control for *Johnny 2*, we replicated Whitten's recruitment poster and matched her pay scale. In retrospect, matching the means of publicity appeared to be irrevellant, matching the pay scale was a good idea, but matching the recruitment poster may not have been.

4.1 The University Population

We originally had some concerns that placing our posters only at MIT would result in a skewed population that would be considerably more educated and computer savvy than the population in general. We found that this was only partially true.

While the majority of our study participants were highly educated—all of our subjects had attended some college, and roughly half were either in or had completed an advanced degree—in the end this highly-educated population didn't seem particularly well-skilled in the use of email security systems. The relatively poor performance that even PhD-level scientists exhibited when attempting to complete the *Johnny 2* protocol may simply reflect the degree of specialization within the technical community: there is no obvious reason that physicists should be any better equipped at dealing with email security than an English literature major.

We also found that the members of the MIT community inter-

msg #	Sender	Content
1	Maria Page	Introductory message introducing Maria and giving the Campaign Coordinator details of the campaign worker's stories. The Coordinator is told to reply. This message provides the subject with information and verifies that they can read and respond to written instructions. This message is also an internal control: Subjects that do not respond to Message #1 within a reasonable amount of time are disqualified and withdrawn from the experiment.
2	Maria Page	The Campaign Schedule and a command telling the Coordinator to send a copy of the schedule to Paul Butler and Dana McIntyre. This message further tests that the subject can respond to a written command from Maria. It also gets the subject into the rhythm of reading an email message and responding by sending out the schedule.
3	Ben Donnelly	Ben asks the Campaign Coordinator for a copy of the schedule.
4	Attacker Paul	Paul says that he is having computer problems and asks the Coordinator to send a copy of the schedule to both Paul's campaign account and his personal Hotmail account, `Paul_J_Butler@Hotmail.com`.
5	Attacker Sarah	Attacker Sarah sends email from her Hotmail account `sara_carson_personal@hotmail.com` saying that she is working at home and asking that the schedule be sent to the personal account.
6	Attacker Maria	If the subject does not succumb to both message #4 and message #5, then message #6 is sent. This message is an unsigned message that purports to come from Maria Page, the Campaign Coordinator's boss. Attacker Maria says that she has tried to call the office but that the phones are not working. Maria says she has been on the phone with both Paul and Sarah and that they both need copies of the schedule; please send them! Now! Do it!
7	Maria Page	In this message, the real Maria Page asks the Campaign Coordinator to send copies of the schedule to Ben Donnelly and Sarah Carson. Some subjects were confused that Maria sent this message, as they had already sent a copy of the schedule to Ben in response to message #3. (Presumably Maria didn't know that Ben had asked for the schedule.) This is a very useful test message to probe precisely what the subject thought had happened in message #6.
8	Maria Page	Maria thanks the subject for participating in the experiment and tells the subject that it is now time for the "Debriefing Interview." Although it wasn't strictly needed, this message gave the experimenter a gentle and in-scenario way to end the experiment.

Table 2: The *Johnny 2* Messages

ested in participating in user studies included a large number of freshmen who had not received much education beyond high school, clerical workers who had not attended MIT as undergraduates, and even some community members who had no significant college education (and certainly no technical education) to speak of.

4.2 Pay scales for human subjects

Because our goal was to match Whitten's experimental procedure as closely as possible, and because Whitten had paid her subjects $20, we decided to do the same. At the time it was pointed out that Whitten paid her subjects for what turned out to be a two-hour test, while the *Johnny 2* test typically took users 30 to 60 minutes. Nevertheless, because whitten prominently advertised that her subjects would be paid "$20," we felt that we had no choice in the matter.

In retrospect, paying subjects $20 turned out to be an excellent strategic move. Most studies at MIT pay their participants between $5 and $15 and have a relatively difficult time getting subjects and a wide range of subjects. Paying $20 we found that we were bombarded with people requesting to be in our study.

Although some may feel that paying subjects $20 to $40 per hour for their time is inappropriate in the university setting, we feel these costs are easier to justify if the subjects are viewed instead as expendible resources that need to be purchased in order to perform the research. A researcher who would be willing to purchase a $500 plane ticket if a $250 ticket is not available should not be concerned about spending $500 instead of $250 to purchase 25 human subjects for an experiment—especially if the additional money spent means that a wider range of subjects will be available and that the experiment can be completed in a shorter period of time. (Concluding all trails within a short period of time helps minimize the impact of external factors on the human subject test.)

4.3 The Computer Security "flag"

A more significant problem with the Whitten recruitment strategy was identifying her study as one involving "computer security." Because of our previously-stated desire to replicate Whitten's protocol, we copied this language. (In fact, we copied Whitten's recruitment poster, changing only her name, phone number, and email address.)

It isn't clear what sort of impact results from identifying the study as a study that is designed to test a computer security tool. This could, for example, cause a selection bias in recruitment. It could cause a performance bias, making subjects more sensitive to potential security issues. Or it could have no impact whatsoever.

It might be interesting for some future study to recruit two populations, one with a "security" poster and one without such a poster, and see if the two populations perform with any significant difference on a standardized task such as *Johnny 2*.

4.4 Screening based on email client

We saw some apparent differences between the manner in which habitual webmail users and users of stand-alone email applications interacted with our Outlook Express email client. Specifically, webmail users would click on the word "inbox" in the Outlook Express user interface to check their mail, rather than using the button labeled "Send/Recv." We discovered that out users who were predominantly Macintosh users had significant problems selecting text with the Microsoft mouse. Finally, we noted that web mail users were confused by instructions on button: they read these instructions as status messages that the computer was using to inform the user, rather than as commands that the user could invoke to tell the computer to do something. Users were also confused between single-clicking and double-clicking.

Our experience in the *Johnny 2* user study indicates that other experimenters using this scenario may wish to control for prior experience with specific styles of email, if not for specific email clients.

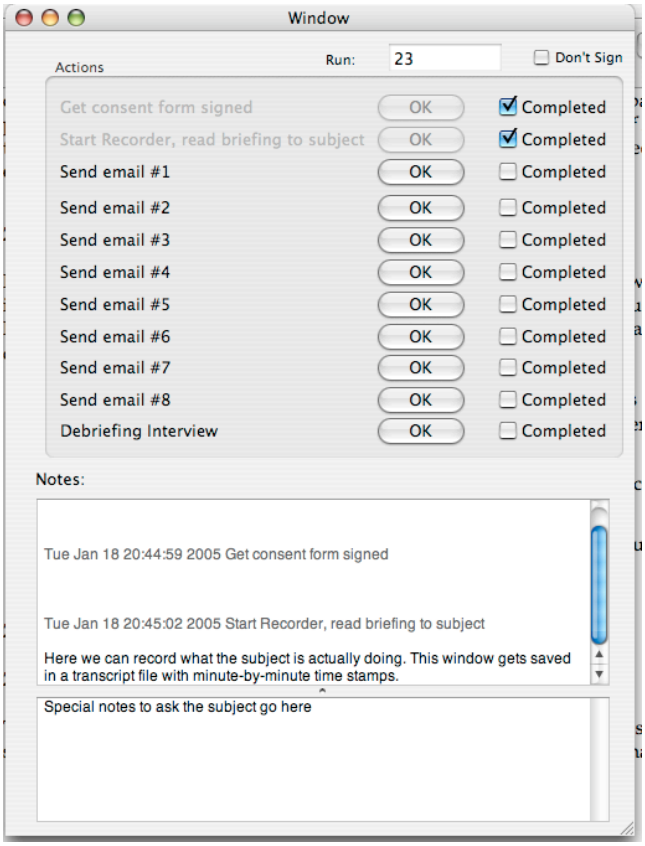

Figure 5: The *Johnny 2* Experimenter's Work Bench. As the experiment takes place, the experimenter successively presses each "OK" button on and takes notes in the two text areas on the bottom. Notes are automatically timestamped every minute and the notes file is saved with every keystroke. Each transcript is stored in a sequentially-numbered RTF file where it may be reviewed by the experimenter or processed with automated tools.

Earn $20 and help make computer security better!

I need people to help me test a computer security program to see how easy it is to use. The test takes about 2 hours, and should be fun to do.

If you are interested and you know how to use email (no knowledge of computer security required), then call Alma Whitten at 268-3060 or email alma@cs.cmu.edu.

Figure 6: Whitten's recruitment poster

5. ADOPTING JOHNNY 2

We have given considerable thought to ways that the *Johnny 2* protocol could be adapted to testing other kinds of tools that provide secure messaging or anti-spoofing functionality.

Applying *Johnny 2* to other secure e-mail systems such as the current version of PGP, Groove, Lotus Notes, or even S/MIME without Key Continuity Management is relatively straightforward: simply follow the protocol, using the secure messaging system in question to send and receive all relevant email messages. In the case of each of these systems the scenario could either have all secure messaging paraphernalia previously set up by Ben Donnelly, the campaign's IT coordinator, or installing the software and/or creating the keys could be made to be part of the scenario as well. These two groups could even be compared.

Likewise, *Johnny 2* could be used to test the security offered by various secure Instant Messaging services by delivering some or all of the messages over various instant messaging channels. Is an attack delivered by IM more or less likely to be successful than an attack delivered by email? How about an attack that is delivered through a message left on a voice-mail system?

The attack described in *Johnny 2* is similar to a so-called "phishing attack." Currently, a number of research group have developed anti-phishing tools (e.g. [2]). *Johnny 2* could be used to test the effectiveness of these tools by comparing the effectiveness of the attack against a group who uses the tools to those who do not.

A number of rights management systems are currently in the marketplace. These systems use cryptography combined with access control lists and/or PKI to give some users access to documents while denying access to others.[6] An open question with these systems is whether or not the interfaces are easy enough to be used on a daily basis, and whether or not an attacker would be able to convince an operator to unlock a managed document. The *Johnny 2* scenario could be used to test such a system by putting the secret schedule into some kind of managed document—for example, a Mmicrosoft Word file.

6. IMPROVING JOHNNY 2

In general we refrained from attempting to make any changes to the *Johnny 2* scenario that were not required for our study. However, in the course of running the study itself, we discovered several ways that the scenario could be improved.

6.1 Secret Selection

In *Johnny* and *Johnny 2* the secret is a schedule of campaign appearances. Several subjects in our user test of 43 participants found that a schedule was not a credible secret: some didn't understand what the motivation would be for keeping appearances secret; others understood the motivation and disagreed with it. Yet others said that the scheduled appearances could easily be changed if they were inadvertantly leaked, and thus it was not very important to protect them.

We suggest that the secret should be changed to something that is more credible. For example, the secret could be a briefing book for an upcoming debate that is scheduled between the two candidates. Alternatively, the secret could be a list of usernames and passwords that the campaign uses to access online resources, or a list of credit card numbers that the campaign uses to pay for campaign events.

6.2 Campaign Personas

We believe that the number of campaign personnas should be increased from 4 to 20, so that the user can be provided with a "campaign address book" that contains many names that will not be otherwise used in the study. We think that increasing the size

of the campaign address book would remove the temptation of the subject to email the secret to everyone in the campaign. It would also eliminate the potential confusion that we observed where the experimental subject is twice asked to send the secret to Ben Donnelly — once by Ben himself, and once by Maria.

6.3 Hotmail vs. other Personal Email Accounts

One failing of our protocol was that all of the attacker email addresses were on the Hotmail service. Hotmail was picked because it is relatively easy to create, use, and manage Hotmail accounts.

While some of our experimental subjects were profoundly suspicious of Hotmail, others were not. Analysis of the data indicated that those who used Hotmail or had friends who used Hotmail were generally not suspicious of the service. People who refrained from webmail were generally more suspicious.

The use of Hotmail exclusively for attacker personnas meant that a simple strategy for success in this study was simply to "never send anything confidential to Hotmail." Several participants in our control group adopted this strategy.

Hotmail is a well-known webmail system. A simple way to control for this "Hotmail effect" would be to remove Hotmail and have the attackers use a different, non-descript domain. Ideally the attackers should have all used different domains for their allegedly "personal" email addresses.

6.4 Political Leanings

Two subjects noted that message #2 indicates that the fictitious candidate apparently is apparently quite left-leaning. Although this wasn't an issue for the *Johnny 2* experiment that we ran, message #2 should probably have been sanitized to be more politically neutral.

6.5 Variable Payments

Finally, the cash payment could be used as a performance incentive. For example, we could have offered a base payment of $20 plus an additional payment of $5 for each correct message that they sent out and a penalty of $5 for each message that they sent out in error. It would be possible to use run two versions of the *Johnny 2* experiment concurrently, one in which the subjects were incentivized, one in which they were not, and compare the results.

7. CONCLUSION AND AVAILABILITY

We have presented *Johnny 2*, a revised and expanded scenario based on the *Johnny* protocol developed by Whitten and Tygar. We believe that other experimenters can benefit by using this and other standardized scenarios—either verbatim or with modification—in conducting tests of security software.

Full source code for the Experimenter's Work Bench, the *Johnny 2* messages, and the scripts that we used to create the *Johnny 2* PKI are being made available on our website at `http://www.simson.net/johnny2/`.

8. REFERENCES

[1] P. A. Carney, A. J. Dietrich, M. S. Eliassen, M. Owen, and L. W. Badger. Recognizing and managing depression in primary care: a standardized patient study. *Journal of Family Practice*, 48:965–972, December 1999.

[2] N. Chou, R. Ledesma, Y. Teraguchi, D. Boneh, and J. C. Mitchell. Client-side defense against web-based identity theft, 2004. URL `http://crypto.stanford.edu/SpoofGuard/webspoof.pdf`.

[3] J. P. Freer and K. L. Zinnerstrom. The palliative medicine extended standardized patient scenario: A preliminary report. *Journal of Palliative Medicine*, 4:49–56, 2001.

[4] S. Garfinkel and R. Miller. Johnny 2: A user test of key continuity management with s/mime and outlook express, 2005. Submitted to Usenix Security 2005.

[5] P. Gutmann. Why isn't the Internet secure yet, dammit. In *AusCERT Asia Pacific Information Technology Security Conference 2004; Computer Security: Are we there yet?* AusCERT, May 2004. URL `http://www.cs.auckland.ac.nz/~pgut001/pubs/dammit.pdf`.

[6] A. Mehta. Information rights management in office 2003. *TechNet*, Winter 2005. URL `www.microsoft.com/technet/technetmag/issues/2005/01/OfficeSpace`.

[7] K. D. Mitnick and W. L. Simon. *The Art of Deception*. John Wiley & Sons, 2002.

[8] I. B. Myers and P. B. Myers. *Gifts Differing: Understanding Personality Type*. Davies-Black Publishing, May 1995.

[9] S. Schweitzer. Parties call foul over N. H. phone-jaming suit. *The Boston Globe*, October 23 2004.

[10] TechSmith. Camtasia studio 2.1, 2005. URL `http://www.techsmith.com/products/studio/`.

[11] A. Whitten. *Making Security Usable*. PhD thesis, School of Computer Science, Carnegie Mellon University, 2004.

[12] A. Whitten. Personal communication, December 6 2004.

[13] A. Whitten and J. D. Tygar. Usability of security: A case study. Technical report, Carnegie Mellon University, December 1998. URL `citeseer.ist.psu.edu/whitten98usability.html`.

[14] A. Whitten and J. D. Tygar. Why Johnny can't encrypt: A usability evaluation of PGP 5.0. In *8th USENIX Security Symposium*, pages 169 – 184. Usenix, 1999. URL `citeseer.nj.nec.com/whitten99why.html`.

Acknowledgments

The authors wish to express their thanks to Alma Whitten for answering so many questions regarding the original *Johnny* experiment. Rob Tannen graciously reviewed an earlier version of this paper and offered useful comments.

APPENDIX

A. THE JOHNNY 2 MESSAGES

A.1 Message #1

From: Maria Page <mpage@campaign.ex.com>
To: Campaign Coordinator <ccord@campaign.ex.com>
Subject: Welcome to the Campaign!
Text: Dear Campaign Coordinator,

Please click "reply" and send me a brief email message when you read this to let me know you are ready.

Hi there! Once again, I wanted to thank you for taking time out of your busy schedule to work with us here on the Senator's re-election campaign. It's just a few weeks to go before the election and we really, really, *really* can use your help!

I've cc'ed the other team members on this email. They are:

- **Paul Butler** `butler@campaign.ex.com`, our campaign finance manager and chief election strategiest.
- **Ben Donnelly** `bend@campaign.ex.com`, who is officially Paul's assistant, but who also runs the IT for our campaign.

Ben's also a full-time student at the University of Pennsylvania.

- **Sarah Carson** `carson@campaign.ex.com`, who is a full-time graphics designer. She designed that slick bumper sticker that is on the back of your car! She also does all of our press releases.
- **Dana McIntyre** `dmi@campaign.ex.com`, who is our office manager. Normally Dana would be there with you in the office, but she's out this week because her husband is having surgery! (Don't worry, it's a routine procedure.)

Because Dana is out of the office this week, we're going to be relying on you to help out in a big way! Don't be nervous, but we are counting on you!

Please click "reply" and send me a brief email message when you read this to let me know you are ready.

—Maria

Comment: This is the initial message from Maria to the Campaign Coordinator. The message displays as yellow because it is the first message received from the email address mpage@campaign.ex.com. Maria cc's the other campaign members on the email — Paul Butler, Ben Donnelly, Sarah Carson, and Dana McIntyre.

A.2 Message #2

From: Maria Page <mpage@campaign.ex.com>
To: Campaign Coordinator <ccord@campaign.ex.com>
Subject: Speaking dates for Pennsylvania
Text: Dear Campaign Coordinator,

Thanks for your email. It's great that you are settling in. There is chocolate in the file cabinet on your left if you want any. Also, feel free to use the phone for phone calls, but *be sure that at least one phone line is available at all times*.

In any event, I want you to know that we have finalized the speaking dates for Pennsylvania. Here they are:

- Monday 10/10 Harrisburg
  - 9:30am - Rally on the Green. Lots of media attention.
  - noon - Photo-op at city library.
  - 3:30 - Sit-in at the mayor's office.
- Saturday 10/15 Hershey
  - 10:00am - chocolate factory tour.
  - 6:00pm - campaign dinner to honor chocolate workers.
- Tuesday 10/18 Philadelphia
  - 10:00am - "Break the bell" at the Liberty Bell.
  - 4:00pm - Constitution 2 at Liberty Hall.
- Friday 10/21 Pittsburgh
  - 10:00am - Toxic workshop at Pittsburgh Airport.
  - 2:00pm - Meet the workers at the docks.

It's important that we get this information out to the other members of the campaign. **But we are not releasing this information to the public until the day of each event.** If the opposing campaign discovers our schedule, they will arrange to have protesters show up at our events! That would be *really, really bad*.

Indeed, the other campaign may be trying to steal this information!

I'm having a problem with my email right now.

Please send the schedule to Paul Butler `butler@campaign.ex.com` and Dana McIntyre `dmi@campaign.ex.com`. Thanks!

Remember, if *anybody on our team* asks for a copy of the schedule, please send it out to them! But please don't send it to anyone else.

—Maria

Comment: This is the second message from Maria to the Campaign Coordinator. In this message Maria asks that the Campaign Coordinator send the secret to butler@campaign.ex.com and to demi@campaign.ex.com. In our test, this message contains the "secret" that must be distributed to the other campaign members and simultaneously shielded from the attackers. This arose confusion on the part of some experimental subjects, who wanted to know why Maria couldn't just email the secret herself. In a revised Johnny 2 *protocol, we suggest that the secret be a Microsoft Word file that is on the computer's desktop.*

A.3 Message #3

From: Ben Donnelly <bend@campaign.ex.com>
To: Campaign Coordinator <ccord@campaign.ex.com>
Subject: I need a copy of the Pennsylvania dates!
Text: Dear Campaign Coordinator,

Hi! This is Ben Donnelly. I run the computer systems for the campaign. I'm also a full-time student at Penn. Welcome on board!

I just got off the phone with Maria. She said that you have a copy of the speaking dates for Pennsylvania and that you could email them to me.

Can you please email me the schedule? I'm trying to make sure that we will be able to coordinate wireless Internet coverage at each of the stops.

Thanks.

—bend

Comment: This is the first message from Ben Donnelly. Since it really is from Ben, the Campaign Coordinator should send the secret.

A.4 Message #4

From: Paul J. Butler <butler@campaign.ex.com>
To: Campaign Coordinator <ccord@campaign.ex.com>
Subject: Something is wrong with my email!
Text: Dear Campaign Coordinator,

Did you get my previous email? Something screwy is going on. I sent you a long message and it bounced... Did you get it?

Anyway, it's **urgent** that I get a copy of the Candidate's schedule within the next half-hour — I'm about to sign a deal with a major outdoor advertising company.

I need you to send me a copy of the candidate's schedule to **both** this account **and** my Hotmail account? You can find the address in the campaign phone book — use `Paul_J_Butler@Hotmail.com`.

Thanks!

Comment: This is the first attack message. This is a spoof message that could easily be sent by an attacker. The Campaign Coordinator should not follow the instructions in Message #4 because it does not come from a trusted source.

Some subjects were confused by this message. One subject didn't understand why the campaign was trying to sign an outside advertising contract to publicize a schedule that is being kept secret. (Presumably, the subject didn't realize that it's reasonable to purchase outdoor advertising space in advance at locations of planned rallies — both to get the coverage and to prevent the opposing campaign from purchasing the space for attack advertisements.) Another subject didn't understand why there would be a rush to purchase a contract for a campaign rally that was scheduled for many months in the future. Both of these points of confusion could be addressed by using secret that is different from a list of campaign stops.

A.5 Message #5

From: Sarah Carson <sara_carson_personal@hotmail.com>
To: Campaign Coordinator <ccord@campaign.ex.com>
Subject: Dates for Pennsylvania?
Text: Dear Campaign Coordinator,

Hi there! I'm working from home this week and can't access my email from work, so I'm using HotMail.

I'm putting together the art for the Pennsylvania events. I need dates! Can you please send them to my HotMail account? It's `sara_carson_personal@hotmail.com`.

I'm using HotMail to send this message, so you can probably just hit "reply."

Thanks so much. I really appreciate this.

—sc

Comment: This is second attack message. In this escalation of the attack, the attacker has created a new HotMail identity that has a name similar to Sarah Carson's (although the Hotmail account was inadvertently misspelled, an inconsistency is completely plausible in the case of an attack and that, in fact, was only noticed by two of our experimental subjects.) Many subjects fell for this ruse.

A.6 Message #6

From: Maria Carson <mpage@campaign.ex.com>
To: Campaign Coordinator <ccord@campaign.ex.com>
Subject: Please send the schedule to Butler and Sarah!
Text: Dear Campaign Coordinator,

Hi there! I'm working from home this week and can't access my email from work, so I'm using HotMail.

I'm putting together the art for the Pennsylvania events. I need dates! Can you please send them to my HotMail account? It's `sara_carson_personal@hotmail.com`.

I'm using HotMail to send this message, so you can probably just hit "reply."

Thanks so much. I really appreciate this.

—sc

Comment: This is the third attack message. In this message, the attacker has forged a message from the mpage email address which tells the Campaign Coordinator to follow the instructions in the previous attack messages. This style of attack, which involves using multiple personae, is in the style of the attacks described by Mitnick [7].

A.7 Message #7

From: Maria Carson <mpage@campaign.ex.com>
To: Campaign Coordinator <ccord@campaign.ex.com>
Subject: Please send the schedule to Ben and Sarah
Text: Dear Campaign Coordinator,

Hi once again! We're going to be wrapping things up here pretty soon. You've been really great so far.

Can you please send a copy of the schedule to **Ben Donnelly** (`bend@campaign.ex.com`) and to **Sarah Carson** (`carson@campaign.ex.com`)?

Thanks!

—Maria

Comment: This message is the third legitimate message sent by Maria Carson to the Campaign Coordinator. In it, Carson asks the Coordinator to send the schedule to Sarah Carson, the one Campaign volunteer who has not legitimately received the schedule.

A.8 Message #8

From: Maria Carson <mpage@campaign.ex.com>
To: Campaign Coordinator <ccord@campaign.ex.com>
Subject: One last thing...
Text: Dear Campaign Coordinator,

Thanks so much for all of your help today. It's now time for the Debriefing Interview!

—Maria

Comment: This message is the fourth legitimate message sent by Maria Carson to the Campaign Coordinator. It informs the test subject that the test is over.