Class Notes: February 2, 2006
Topic: User Testing II
Lecturer: Jeremy Hyland
Scribe: Rachel Shipman

Why Can't Johnny Encrypt? A Usability Evaluation of PGP 5.0

Alma Whitten and J.D. Tygar

This article has three main points. First, effective security requires a different usability standard than other types of consumer software. Second, security mechanisms are effective only when used correctly. Third, it attempts to outline a specific definition of usability for security.

"Security mechanisms are effective only when used correctly" is the most important point since risk of abandonment or misunderstanding on the part of the user is the largest threat to the successful application of security mechanisms.

Note: The article name comes from a famous journalism piece on literacy called "Why Can't Johnny Read?"

About the authors: Whitten is a CMU grad who now works are Google, Tygar was once a professor at CMU.

To illustrate, the authors designed a usability study of PGP 5.0 for the following reasons:

- PGP has a well designed user interface by consumer software standards
- Marketed as being developed with a focus on GUI design
- Marketed as making cryptography accessible for novice computer users

Their usability study employed the following techniques:

- Cognitive Walkthrough
- Laboratory User Test

Whitten and Tygar's definition of usability for security:
Security software is usable if the people who are expected to use it:

- Are reliably made aware of the security tasks they need to perform
- Are able to figure out how to successfully perform those tasks
- Don't make dangerous errors
- Are sufficiently comfortable with the interface to continue using it

The authors identify 5 problematic properties of security relating to user interface design:

- The unmotivated user property
    - Security is a secondary goal
- The abstraction property

    - Security policies are generally systems of abstract rules for deciding whether to grant access to resources
- The lack of feedback property
    - Providing good feedback for security management is a difficult problem
- The barn door property
    - Once a secret has been left accidentally unprotected there is no way to be sure that it has not already been unread by an attacker.
- The weakest link property
    - The security of a networked computer is only as strong as it's weakest component.

Using the general definition of usability for security, the authors defined this focus for their evaluation: **"If an average user of email feels the need for privacy and authentication, and acquires PGP with that purpose in mind, will PGP's current design allow that person to realize what needs to be done, figure out how to do it, and avoid dangerous errors, without becoming so frustrated that he decides to give up on using PGP after all?"**

There was some skepticism about the focus statement:

- Did the authors predict the outcome of the user test and write the focus statement with this result in mind?
- How do you know if an "average user" feels the need for privacy and authentication?
- The average user was not defined by the article, and if the scenario forces the users into PGP how do you know if they would find the need for this application (or similar) on their own?

To evaluate this focus, the authors started with a cognitive walkthrough of the PGP 5.0 interface. They also employed aspects of heuristic evaluation, another usability technique based on adherence to high-priority usability principles.

Issues identified using cognitive walkthrough:

- Visual metaphors are heavily used in PGP 5.0. Keys, locks, and pens are heavily used to represent the concepts of public and private keys as well as digital signatures. However, these metaphors break down since they do not maintain continuity with the real world.
- Different Key Types to indicate different encryption algorithms
- Invisibility of key server option
- Key management policies names are ambiguous and automatically assigned
- Some user errors are irreversible; deleting the public key, accidentally publicizing a key, accidentally revoking a key, forgetting the pass phrase, and failing to back up the key rings
- Inconsistent use of security terminology

- ♦ Cluttered information screens with unnecessary information confuse users

Summary of the user test:

- ♦ Test scenario: participant is to play a campaign coordinator. They must send email to the campaign team using PGP for privacy and authentication. There were 12 participants and each was given 90 minutes to complete the task. Task completion would require the following tasks:
    - o Generate a key pair
    - o Get the team members public keys
    - o Make their public key available
    - o Type a secret message in an email
    - o Sign the email with their public key
    - o Encrypt using the team members public keys
    - o Send the email
    - o Decrypt a response
    - o Verify the digital signature of the reply
- ♦ Results
    - o 3 sent the email without encryption
    - o 1 user forgot her pass-code, had to create a new key pair
    - o 1 user was unable to encrypt at all
    - o failure to understand the public key model was widespread
    - o only 3 users were able to decrypt the reply message
    - o 8 people got the public keys, but 5 needed help via email
    - o 4 people were able to send an encrypted message
    - o it was unclear if the users knew they were digitally signing or checking a signature upon encrypting or decrypting a message.
    - o Conclusion: PGP 5.0 is not sufficient to make computer security usable for people who are not knowledgeable in security.

Questions raised in class about the validity of the user test:

- ♦ Is the scenario too complex? General consensus was that there should be security software that would allow you to complete the described task in 90 minutes
- ♦ Documentation was provided, but one student made the observation that the version of PGP he used contained all text instructions without step by step instructions with screen shots. The suggestion of a wizard for first time PGP users came up.
- ♦ Who are the target users for this application? Whitten and Tygar used people who were familiar with email, but wouldn't the audience be more specific?

- Is the main problem with the PGP 5.0 interface failure to hire a UI team? General consensus was that a UI team would help PGP define the answers to the following questions:
    - What is the maximum number of things a user has to learn to effectively use the software?
    - How much of this could be done in the background?
    - What does the user absolutely have to be aware of (i.e. is the digital signature valid?)
- Lab studies generally try to decrease the number of variables, but real use context is full of variables that were not considered in this study like training and help from co-workers
- One student observed that although some of these problems were fixed in the more recent versions of PGP 5.0, the addition of features has added to the complexity of the interface
- Ability to email questions and get varied responses depending on who answered the question damaged the ability to recreate the study to verify results

Johnny 2

Garfinkel and Miller, 2005

The main idea of this article is that the key problem with PGP 5.0 was the key model it employed. They theorized that by replacing this certification model with Key Continuity Management (KCM), novice users would be able to easily encrypt and decrypt messages.

They tested the effectiveness of a prototype KCM system using laboratory user tests as the usability evaluation method. They also added another dimension to this user test: the presence of attackers. The attackers would employ one of the three following tactics to gain access to the secret content of email messages:

1. New key attacks
2. Unencrypted message attacks
3. New identity attacks

Users were allowed to email questions like in Johnny 1, but this time email responses were selected from a standardized list. In addition, the questions and responses were tracked as part of the study.

The KCM prototype has the following features; tight email client integration, visual cues in the form of border colors that alter users to encryption status, and basic automated key distribution.

The participants in the user study were broken into one of three test circumstances:

1. No KCM
2. Color

3. Color and briefing

Results:

- No users had difficulty with the basic encryption/decryption process
- Few users understood signature integrity guarantees
- Authenticity of the attack message was difficult for users to identify
- No user was able to avoid attacks 100% of the time
- Group 1 had the hardest time resisting new key attacks and unencrypted message attacks
- All groups had trouble with the new identity attack
- Users has misconceptions about the security of sealed messages
- Addressing surface interface issues helped users in the scenario. For example, the "do not trust" button was easy for user to understand.
- Vastly simplifies email encryption/decryption compared to PGP 5.0

In class discussion brought up the following:

- Use of yellow could be confusing – there are cultural connections between yellow and warning, although it is not being used that way in this application
- The user studies are testing very different technologies under very different circumstances – should it really be called Johnny 2?
- Who uses email encryption?
    - Dissident groups, generally funded by non-profit organizations
        - Risk to being identified because content of message is unacceptable to government/leadership
    - Some security experts think everyone should use it all the time
    - Activist organizations
    - "bad people"
    - Used to send confidential information within a company. Example: performance evaluations at technical research and development company where anyone could hack into the email system.
- Is PGP crack-able?
    - Chance for cracking PGP algorithm is lowered by regular change in encryption algorithm employed

Class exercise:

Divide into teams of 3-4 and create a user study for one of the following objects:

Cell phone
Calculator
Portable CD Player

Design the use scenario and use a member of another group as the participant.
Think about:

- How well does the interface support the participant?
- How will you measure success/failure of different aspects of the task?
- Does this help you generate ideas for redesign of the object?