NOKIA

Research Center
NRC-TR-2007-002

Usability Analysis of Secure Pairing Methods

Ersin Uzun[1,3], Kristiina Karvonen[2], N. Asokan[2,3]
[1] University of California, Irvine, USA
[2] Helsinki University of Technology, Helsinki, Finland
[3] Nokia Research Center, Helsinki, Finland
http://research.nokia.com
January 10, 2007

Abstract:
Setting up security associations between end-user devices is a challenging task when it needs to be done by ordinary users. The increasing popularity of powerful personal electronics with wireless communication abilities has made the problem more urgent than ever before. During the last few years, several solutions have appeared in the research literature. Several standardization bodies have also been working on improved setup procedures. All these protocols provide certain level of security, but several new questions arise, such as "how to implement this protocol so that it is easy to use?" and "is it still secure when used by a non-technical person?" In this paper, we attempt to answer these questions by carrying out a comparative usability evaluation of selected methods to derive some insights into the usability and security of these methods as well as strategies for implementing them.

Index Terms:
Usable Security
Secure Pairing
User Interaction
Secure Pairing Standards

Usability Analysis of Secure Pairing Methods[†]

Ersin Uzun[1,3], Kristiina Karvonen[2], N. Asokan[2,3]

[1] University of California, Irvine, USA
[2] Helsinki University of Technology, Helsinki, Finland
[3] Nokia Research Center, Helsinki, Finland

euzun@ics.uci.edu, kk@tml.hut.fi, n.asokan@nokia.com

Abstract. Setting up security associations between end-user devices is a challenging task when it needs to be done by ordinary users. The increasing popularity of powerful personal electronics with wireless communication abilities has made the problem more urgent than ever before. During the last few years, several solutions have appeared in the research literature. Several standardization bodies have also been working on improved setup procedures. All these protocols provide certain level of security, but several new questions arise, such as "how to implement this protocol so that it is easy to use?" and "is it still secure when used by a non-technical person?" In this paper, we attempt to answer these questions by carrying out a *comparative* usability evaluation of selected methods to derive some insights into the usability and security of these methods as well as strategies for implementing them.

1 Introduction

The process of setting up a security association between two devices is sometimes referred to as *pairing.* Secure pairing of electronic devices that lack any previous association or infrastructure support, is a challenging problem especially when it needs to be done by ordinary end users without technical expertise. The increasing popularity of powerful mobile electronic devices has made this problem more urgent than ever. Laptops, personal digital assistants (PDAs) and mobile phones all have integrated advanced communication technologies. When the same devices are used for monetary transactions also, the security of these protocols gains a whole new importance. However, no standard user friendly method for establishing secure communication among arbitrary devices exists.

Recently, several different proposed solutions to this secure device pairing problem have appeared in the research literature. Typically these protocols utilize human authenticated and possibly location-limited [6] auxiliary communication channels including visual [1,2], aural [3], short-range wireless channels like Near Field Communications (NFC) [8], and actual physical contact. Each of these proposals makes its own assumptions about the hardware capabilities of devices involved.

[†] A short version of this report appears in the proceedings of Usable Security 2007 workshop.

Several standardization bodies also recognized the seriousness of the problem and have begun work on specifying more usable and more secure procedures for device pairing. Wi-Fi Alliance is working on specifications for *Wi-Fi Protected Setup* [9]. Microsoft has released specifications for *Windows Connect Now-NET* [12], which is closely related to Wi-Fi Protected Setup. Bluetooth Special Interest Group has released a white paper on *Simple Pairing* [10] and is expected to release the specifications soon. The Universal Serial Bus (USB) forum has recently released the specifications for *Wireless USB Association Models* [11] which specifies the procedures for pairing two Wireless USB devices. Unlike the research papers, the standards specifications have to consider devices with a range of hardware capabilities. Consequently, the specifications do not dictate a single pairing method. All of them support the use of at least one type of secure auxiliary channel. For example, Bluetooth Simple Pairing supports the use of NFC and Wireless USB Association Models support the use of USB cables. All the specifications also allow the users themselves to be used as auxiliary channels (See Section 3).

To the best of our knowledge, no comparative usability study of user interaction methods for secure pairing exists. We conducted a comparative usability analysis of different methods in order to identify user preferences, evaluate usability as well as to infer general guidelines for implementing some of the proposed pairing methods.

A single test user cannot effectively compare more than a handful of pairing methods in one test session. Therefore, in our study we concentrated on those user interaction methods implied by the emerging standards specifications. Based on a first round of testing, we refined and narrowed the tested interaction methods further and carried out a second round of testing.

The rest of the paper is as follows. In Section 2 we briefly review other work investigating the usability of pairing methods. In Section 3, we describe the user interaction methods we selected based on the emerging standards specifications. In section 4, the study is explained in more detail and its results are discussed in section 5. We finalize the paper by giving our future work plans in section 6.

2 Related Work

Although a number of papers have proposed different solutions to the secure device pairing problem, most of them did not report on any significant usability testing. One exception is the Network-in-a-Box project by PARC [5]. They use location limited channels (such as infra-red, physical contact, USB-storage) to provide human verifiable authentication of devices as a pre-requisite to admitting them to a wireless network. Their user testing was to compare the usability of the proposed approach with the traditional methods for configuring wireless network clients. In contrast, our objective is to compare the usability of different proposed approaches to one another.

3 Pairing protocols and user interaction methods

Based on the pairing protocols described in the emerging specifications for secure device pairing [10,12,9,11], we initially selected five different user interaction methods to be tested, as described below.

In all the emerging specifications, the typical approach for secure pairing consists of running Diffie-Hellman key agreement protocol over the insecure channel between the devices and then authenticating this key agreement. Authentication is achieved by transferring some information via a secure auxiliary channel. In this paper, we focus on the case where the users themselves constitute the secure auxiliary channel. The auxiliary channel is used chiefly in one of two ways:

A. Transfer short string(s) so that integrity checksums computed independently by either device can be compared.
B. Transfer a short secret passcode so that both devices share the same short secret.

In approach A, both devices execute a *short authenticated string (SAS) protocol*, such as those described in [15,14,4]. Each device then independently computes a short checksum based on its view of the protocol run. The SAS protocols ensure that if there is an *active* man-in-the-middle, the two checksums are likely to be different. Bluetooth Simple Pairing specification [10] and WUSB Association Models specification [11] support this approach to secure device pairing. The former requires 6 digit checksums while the latter requires 2-4 digit checksums. Neither explicitly specifies the user interaction by which the checksums are compared. There are three obvious possibilities for the interaction methods:

1. **Compare-and-Confirm**: Each device shows its checksum on its display. The user is then prompted to compare the displayed strings and indicate, on each device, whether the two strings are the same or not.
2. **Select-and-Confirm**: During standardization discussions, there was some concern that the *Compare-and-Confirm* method might be too easy for the users leading to their answering the prompt without actually doing the comparison. A comparison method that forces the user to pay more attention might be preferable. In the *Select-and-Confirm* method, one device shows the checksum on its display. The other device shows a *set* of values including its own checksum, as well as some other randomly chosen strings. On the second device, the user is asked to select the entry from the set that matches the string shown on the first device, or indicate a failure if there is no matching value. If the entry chosen by the user matches its own checksum, the second device indicates success. Otherwise it indicates a mismatch. On the first device, the user is prompted whether the second device indicated success or not.
3. **Copy-and-confirm**: Not all devices have displays. A typical pairing scenario is between a phone/computer and a keyboard. The *Copy-and-Confirm* method is intended to be used in such scenarios. The device with the display shows its checksum and asks the user to type this value into the second device. The second device compares the entered value with its own checksums and indicates success if the values are the same. On the first device the user is prompted whether the second device indicated success or not.

In approach B, both devices execute a *short-secret authentication* protocol. Both WiFi Protected Setup [12] and Bluetooth Simple Pairing [10] take the approach of splitting

the shared secret into k ($k > 1$) equal-sized components and running the MANA III protocol [13] k times where in each round each party demonstrates its knowledge of the k^{th} component. WiFi Protected Setup uses 2 rounds and requires a 4 or 8 digit passkey. Bluetooth Simple Pairing uses 20 rounds and requires a 6 digit passkey. In both cases, the passkey should not be used more than once. Unlike the checksum in approach A, the passkey must be kept secret from attackers until the pairing process has successfully completed. There are two possible user interaction methods:

4. **Copy**: One device chooses a passkey and displays it to the user and the user is asked to type the displayed value into the second device. The devices automatically run shared secret authentication protocol which succeeds or fails depending on the user's ability to copy the passkey correctly into the second device and the presence of an active attacker. Unlike in the *Compare-and-Confirm* method, no further user interaction is needed here.
5. **Choose-and-enter**: The user is asked to choose a random passkey and enter it into both devices. Then the devices automatically run shared secret authentication protocol which succeeds or fails depending on the user's ability to enter identical values into both devices and the presence of an active attacker.

In all of the above approaches, the likelihood of a successful man-in-the-middle attack is inversely proportional to the size of the set of values the passkey or checksum can take [21]. The only exception is WiFi Protected Setup, where the level of security is inversely proportional to *half* the length of the passkey space. In other words, to achieve a 4-digit level of security in WiFi Protected Setup, 8 digit passkeys need to be used. The security of all of the approaches is predicated on the assumption that the software implementing the pairing procedure on each device has a *trusted path* to the user: approach A requires that the attacker cannot hide or alter the UI (messages and prompts shown to the user) of the pairing procedure on either device; approach B requires further that the attacker cannot read the passcode displayed to the user.

4 The Study

In computer security, even one user error can be too much. In this regard, the principles of usability of security clearly deviate from the general usability principles. Usually, a trial-and-error approach is acceptable for the learning period when taking a new system into use or playing around with the advanced features of, say, an Office application. However, in usability of security this is not possible. The same holds, of course, for other security-critical systems, such as airplane cockpits or management of nuclear power supplies.

In a security-related interaction, we can group user errors into two categories. A *fatal error* results in the violation of a security goal. All other errors are *safe errors*. Although acceptable fatal error rate may change depending on the application, we assume that any non-zero fatal error rate in the sample size of 40 is unacceptable for security applications. With respect to the pairing methods described in Section 3, we consider the following fatal errors in our study. In approach A, a fatal error occurs when the checksums computed by each device are different, but user input causes one or both devices to conclude that the checksums match. Fatal errors are possible in all

three interaction methods of approach A. In approach B, a fatal error occurs if the user chooses an easy-to-guess passkey in the *Choose-and-Enter* method. There is no possibility of a fatal error in the *Copy* method.

This leads to the first two research questions we want to investigate regarding the *security* of the tested methods:

1. Do users accidentally/carelessly make fatal errors in the tested methods?
2. Does *Select-and-confirm* have a lower fatal error rate than *Compare-and-Confirm*?

In addition to the security implications of the interaction methods, we also want to find out the *effectiveness* of the methods both quantitatively, and in terms of user perception. This leads to the next two research questions:

3. How do the methods compare in terms of user perception?
4. How do the methods compare in terms of measurable parameters of effectiveness? (time to completion and total user error rate)

4.1 Test Design & procedure

Introduction of the tests to users: When test users know that they are testing something related to security, their behavior tends to change drastically [7]. In order to keep user behavior realistic, we designed all test material and procedures so that (a) until the end of the test, security-relevance of the procedure is not emphasized, and (b) the feedback on user actions was independent of whether the action constituted a user error or not.

Choice of devices: The test scenario was one user pairing two devices of the same kind. The same user controlling both devices is the most common real-life scenario. But the devices involved are usually not similar. In order to account for this, we used only the most basic user interactions in designing the user interfaces. Similar user interfaces can be implemented in most types of devices.

Test procedure: Users were first given brief introduction to the study. They were then asked to fill out the background questionnaire (Appendix C) to get demographic information and learn about their mobile device usage history. Next, users were given a brief introduction to the devices to show them the basic operations needed during the test, such as how to move the cursor, erasing a character, etc. The tests were then presented to the user sequentially in random order. Finally they filled out the post-test questionnaire (Appendix D). In the post-test questionnaire, users were given screenshots of each tested method for easy reference. They were asked to associate given adjectives (e.g., “easy”, “professional” etc.) with the methods, which method they would like for their own device and what they found difficult about the interactions/UIs during the test. Tests were run in a private room with no disturbance during the whole process. The testing time was around 20 minutes per user including at least 5 minutes of free discussion at the end where they could give us any additional verbal feedback. The testing procedure remained same throughout the study although the tested method variants and test devices changed.

4.2 Test implementation

To investigate the likelihood of fatal errors in the methods involving comparing checksums, we needed to simulate a man-in-the-middle scenario by having the devices use different checksums. To measure effectiveness parameters, we needed to record the time for completion. Finally, we needed to present the tests in random order to account for learning effects. We designed a software framework that aids in all of the above. The framework sets up a communication channel between the two devices for co-ordination and control. It takes care of measuring completion times and logging user actions. It also enables partially automated test planning. All common functionality, such as inter-device communication or logging, is exposed via a simple application programming interface. In effect, the framework allows usability testing of any multi-device distributed application. The test developer needs to implement the graphical user interface, and few service calls which can be invoked by the framework. We intend to publish this framework as open-source in the near future. Further details about this framework can be found in [16].

4.3 Participant profile

We did two rounds of usability tests with 40 participants in each. Both tests were conducted in university environments in two different countries, representing over 10 nationalities, with a clear majority being U.S. and Finnish citizens. We used similar means of recruitment announcement, such as mailing lists and bulletin boards, to attract similar participant groups in both environments. The distribution of gender, age and education of the test participants are given in Table 1.

	First Group (40 people)	Second group (40 people)
Gender	Male: 60% Female: 40%	Male: 70% Female: 30%
Age	18-24: 22% 25-29: 52% 30-34: 15% 35-39: 8% 40+ : 3%	18-24: 20% 25-29: 47% 30-34: 15% 35-39: 5% 40+ : 13%
Education	High School: 13% Bachelor : 30% Masters : 47% Doctorate : 10%	High School: 32% Bachelor : 28% Masters : 25% Doctorate : 15%

Table 1. Participant Profile

The groups had other similar characteristics. In both groups, the average computer usage history for participants was around 12 years and the average computer usage was 7 hours per day. All participants in our study had either a PDA or a mobile phone, or both.

4.4 First Round

In this first round of our study, we conducted our usability tests in a university in the United States.

4.4.1 Material We used iPAQ devices running Windows CE operating system. User interaction consisted of using an on-screen keyboard on a 2.26 x 3.02 inches size color screen. The Windows CE environment is intended for mobile devices but it provides a windowed GUI environment that is similar to PC and other PDA operating systems.

4.4.2 Tested methods Each method described in Section 3 was tested. The settings used for each method are described below (Screenshots can be found in appendix A).

1. *Compare-and-Confirm*: We used randomly generated 4-digit numbers to be presented as "checksums". In half the cases, chosen randomly, we showed different values on the two devices. The issuer prompt was "Check if both devices display the same value". Users were given two button choices labeled as YES and NO to give their answers.
2. *Select-and-Confirm*: The first prompt on the first device was "Please select "XXXX" from the list on the other device" followed by the question "Did the other device indicate success?" and YES/NO buttons. The second device simultaneously showed the instruction "Please choose the value other device is displaying" and a list consisting of four 4-digit numbers, including the value shown on the first device. A success or failure pop-up screen appeared depending on whether the user chose the correct value in the list or not.
3. *Copy-and-Confirm*: The first device showed the text "Enter the displayed key to the other device" followed by a 4-digit checksum and the question "Has the other device indicated success?" The second device instructed the user "Please enter the value the other device is displaying" and showed a success or failure pop-up depending on whether the value was copied correctly.
4. *Copy*: We tested two variants: one using 4-digit passcodes and the other using 8-digit passcodes. The first device showed a key and the text "Enter the displayed key to the other device". The second device instructed the user "Please enter the value the other device is displaying".
5. *Choose-and-Enter*: The prompt was "Choose a 4-digit hard to guess number and enter it into both devices".

4.4.3 Results The data collected is summarized in Table 2.

Method	Variant	Average Completion Time (sec.)	Fatal Error Rate	Total User Error Rate
Compare-and-Confirm		15.6	20%	20%
Select-and-Confirm		22.5	12.5%	20%
Copy-and-Confirm		27.6	10%	20%
Copy	4-digits	20.8	N/A	7.5%
	8-digits	31.7	N/A	5%
Choose-and-Enter		32.7	>42.5%	45%

Table 2. Summary of first round usability tests

Participants were asked to associate given adjectives with the methods. The Participant opinions are summarized in Figure 1. The graph shows the percentage of the participants who associated certain adjective with a certain method variant.

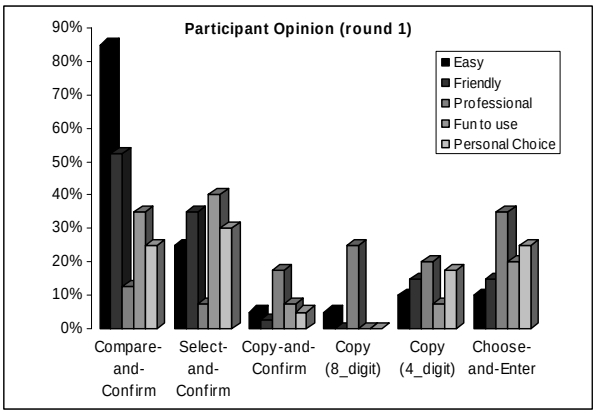


Fig. 1. Summary of participant opinions in the first round.

We can make the following observations

- *Copy-and-Confirm* as well as *Copy* with 8 digit passkeys were perceived to be hard to use.
- Fatal error rate was unacceptably high in all the methods except *Copy*.
- *Select-and-Confirm* had a 7.5% lower fatal error rate than *Compare-and-Confirm*.

The *Choose-and-Enter* method had an extremely high fatal error rate: 42.5% of the users chose passkeys that were in a small set of predictable sequences we screened for. It also had the longest average completion time. Since there is no way to improve the fatal error rate in this method, we decided to abandon it.

The *Copy-and-Confirm* method had a high fatal error rate, and was not perceived to be easy to use. From user feedback, it was evident that users were confused about having to do two things (type a passkey, *and* confirm). Therefore, we decided to abandon this method as well. It implied that in situations where *Copy-and-Confirm* would have been applicable, it would be necessary to use the *Copy* method.

The *Compare-and-Confirm* and the *Select-and-Confirm* methods both had unacceptably high fatal error rates. In *Compare-and-Confirm,* all user error was fatal. We decided to experiment further by modifying the UI in these cases.

The *Copy* method was inherently not prone to fatal errors, although users did not perceive it as a user-friendly method.

In the next round, we decided to focus on the methods *Compare-and-Confirm*, *Select-and-Confirm* and *Copy.*

4.5 Round Two

We conducted our second round of tests in a Finnish university. The participant profile was quite similar to our first study as explained in section 4.3.

4.5.1 Material We used Nokia E60 series mobile phones running Symbian S60 3rd edition. We decided to focus on mobile phones because it is likely that mobile phones will be one of the most frequent executers of secure pairing methods.

All test material, including questionnaires and user interfaces were available in both English and Finnish. Participants chose their preferred test language. Finnish tests were conducted by a native Finnish speaker.

4.5.2 Tested methods We implemented the three variants selected at the end of the first round. Based on the first round experience, we made some changes intended to improve usability and security, as described below. All methods are tested with 6-digit numbers, used either as checksum or passcode. We chose this value because it is the longest value mentioned in the standards [10]. Although [9,12] allow 8 digit passcodes, we ruled it out based on the results of the first round, as well as the established cognitive fact that the maximum number of chunks of information that can be kept in working memory is 7 [17]. In the UI, the numeric code was consistently referred to as a PIN, regardless of whether it was used as a passkey or checksum. Screenshots of the implementations can be found in appendix B.

1. *Compare-and-Confirm*: The wording of the question was changed to "Compare the PIN numbers shown on both devices, are they DIFFERENT?" and user was given two choices of SAME and DIFFERENT. The default response key was assigned to the option DIFFERENT, so that accidental or careless user error will no longer be a fatal error (Note also that the default label used exactly the same word as in the question). This was done in order to gain the users' attention. When a difference is suggested, users tend to concentrate more on finding it (e.g., [18]). Further, Hammer et al [19] have shown that (i) people use positive constraints more intuitively, although they fail to use them perfectly and (ii) the use of negative constraints enables a less natural, but potentially more accurate categorization strategy. This meant that in the usual case, the user's thought process has to deal with something akin to double negation: when the number sequences were the same, the response to the prompt is "no", which the user has to mentally map to the key labeled SAME. This design choice could be a potential source of difficulty since it is well known in cognitive psychology that processing of double negation is more complex and thus slower. We tested two variants, one with matching checksums and the other with non-matching checksums.
2. *Select-and-Confirm*: The selection list offered four choices to select from but "No Match" was added as an option to make the action more intuitive when the correct value is not in the list. Design of the selection screen was changed to target more user attention. The first prompt changed to "Please select the PIN below on other device" followed by the checksum in a separate line and the second prompt "Has the other device indicated success after selection?". The pop-up screen showing success or failure was also redesigned to give explicit next action guidance, E.g. "Successful!, please choose YES on the other device to continue". We tested two

variants one in which the set on the second device included the checksum shown on the first device, and the other in which it did not.

3. *Copy*: Screen text in first device was changed to "Please enter the PIN below into the other device" followed by the PIN in a separate line. Second device prompt was "Please enter the PIN other device is displaying and press OK when you are done".

4.5.3 Results The data collected in this round is summarized in Table 3.

Method	Variant	Average Completion Time(sec.)		Fatal Error Rate	Total User Error Rate
		Match	No match		
Compare-and- Confirm	6-digit & new GUI	16.4	13	0%	2.5%
Select-and-Confirm	6-digit & new GUI	16.4	26.4	5%	7,5%
Copy	6-digit	13	N/A	N/A	2.5%

Table 3. Summary of second round usability tests

We also changed some of the adjectives we used in post-test questionnaire aimed towards getting more precise information while still keeping the gathered information comparable between rounds. A graph summarizing the user opinion is in Figure 2.

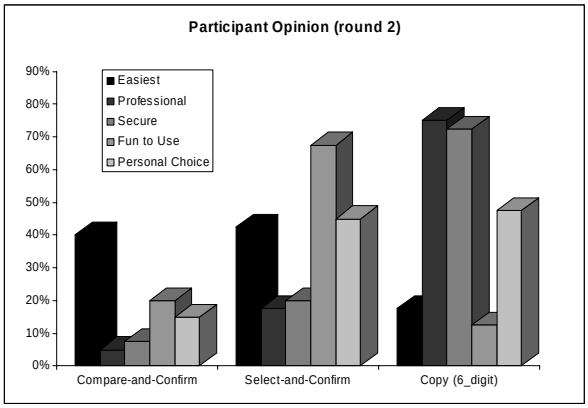


Fig. 2. Summary of participant opinion in second round.

We can make the following observations

- *Compare-and-Confirm* and *Select-and-Confirm* are both perceived as easy but not professional.
- *Compare-and-Confirm* and *Copy* had no fatal errors, while *Select-and-Confirm* still had unacceptable fatal error rate.
- *Copy* is perceived as hard but professional. It was the most preferred personal choice as the pairing method users would like to have available on their devices.

5. Analysis and Discussions

Strictly speaking, the conclusions drawn from the data collected can be considered only as indicative of the whole user base due to the differences in the test set-up between the rounds and the relatively small number of participants.

The fatal error rate in *Compare-and-Confirm* improved significantly from 20% to 0%. There were four differences between the two test rounds: participant groups, devices, checksum lengths, and the UI design. As discussed in Section 4.3, the profiles of the two participant groups were similar. The user interaction is so simple in *Compare-and-Confirm* that the change in devices cannot account for the improvement. The increase in checksum length is not very likely to have *improved* the user error rate; although it cannot be ruled out as a factor since users may have been more careful when faced with a harder task. This leaves us to conclude that the changes to the UI design is the likely cause.

Select-and-Confirm had unacceptable fatal error rates in both rounds. The user actions on the two devices need to be followed strictly in the prescribed order: select on the second device, wait for a response, and only then answer the second prompt on the first device. It is difficult to design the UI so that it strongly guides the user to follow this prescribed order and minimizes the likelihood of flouting it.

The *Copy* method has natural resistance against fatal errors as long as the devices are not compromised or the attacker cannot interfere with the display. The completion time and total user error rate were lower in the second round, which is to be expected since typing digits is easier on cell phones than PDAs.

The *Copy-and-Confirm* and *Choose-and-Enter* methods were abandoned after the first round due to their high fatal error rate and negative user perception. We recommend using *Copy* instead of *Copy-and-Confirm* although *Copy* requires keeping the PIN secret. The C*hoose-and-Enter* method can also be replaced with *Copy* method in many cases.

Users perceived *Compare-and-Confirm* and *Select-and-Confirm* as easy to use, and considered *Copy* difficult. However, they considered *Compare-and-Confirm* and *Select-and-Confirm* to be less secure and less professional than *Copy*. These three properties are often found to be interrelated and also desirable by the users for seemingly irrational reasons (see e.g. [20]).

The popularity of *Compare-and-Confirm* was significantly lower in the second round. This is probably due to the increase in the checksum length, as well as due to the UI change. Some users were surprised by the negative question and unexpected labeling of response actions, and expressed that they would have preferred e.g. the usual "Cancel" and "OK" options instead of "SAME" and "DIFFERENT". User perception may be improved by breaking up the checksum into chunks of two or three digits.

Checksums and passkeys used in the pairing methods are very different from traditional PINs: checksums are not secret; passkeys are limited to single-use and need not be remembered. Nevertheless users assume checksums and passkeys are similar to the type of PINs they are already familiar with. They use this assumption as a reference point for their opinions about tested methods. This had both a negative (PINs are hard to remember) as well as positive (users are familiar with using PINs) bias to the test setting.

Based on these observations we formulate the following guidelines for designing UIs for the tested methods.

- Default user action (e.g., default button) must correspond to the safest choice.
- User actions must be labeled using words that are specific to the task expected from the user. Generic (and familiar) labels like YES/NO, CANCEL/CONTINUE should be avoided. Especially those labels that have direct negative and positive associated meaning should be avoided.
- Multi-step interactions where users can inadvertently and easily change the prescribed order of interactions should be avoided. If such interactions are unavoidable, the UI should make sure that it is difficult to change the prescribed order.

For creating usable procedures with numbers, the cognitive issues involved must be taken into account. For example, checksums and passkeys must not be longer than 7 digits.

Returning to the research questions we started out with in Section 4, we can conclude the following. *Copy* is inherently resistant to fatal errors. Fatal errors in *Compare-and-Confirm* can be avoided by careful design of the UI. *Select-and-Confirm* does not have a lower fatal error rate than *Compare-and-Confirm*. The users clearly differentiated among the methods in terms of ease-of-use and perceived level of security. However the methods tested in the second round were similar in terms of measurable parameters like completion time, fatal and total error rates, and security.

6. Future Work

In this study, we concentrated on obvious interaction models implied by the emerging standards. However, there are other promising methods that either use different auxiliary channels or the human authentication in different means. We already started testing handful of these methods using visual, aural, NFC channels and some methods relying on more basic human sensory capabilities for authentication purposes.

After the first round, we identified several UI improvements. We made *all of them* for the second round for pragmatic reasons. We are currently doing more controlled, smaller-scale tests to better understand the effects of different UI improvements.

We assume throughout the study that the pairing procedure has a trusted path to the user. This can be implemented, for example, if the control of the display cannot be taken out from the pairing software when it is active. When this is not the case, more attack possibilities exist, such as sending a text message to a cell phone during the pairing procedure and hoping that the user will follow instructions in the message. We plan to include these kinds of attack scenarios in our future work.

We plan to make several changes to the test framework to make the tests more realistic. First, people are usually under stress of accomplishing another main task (such as printing a document, transferring a file etc.) while pairing their devices. We plan to change the test framework so that pairing is carried out as a necessary step in a given task involving the transfer of sensitive information, such as synchronizing address books between two devices. Second, the presence of an observer may have had an effect on user behavior. We plan to investigate ways of allowing the user to do

the tests, at their leisure, in a familiar environment without the presence of an observer. This would also allow us to investigate user behavior over time: for example, will the effect of unfamiliar labels may wear off over time.

Acknowledgements: We want to thank several colleagues for their help and feedback. Philip Ginzboorg, Kari Kostiainen, Kaisa Nyberg, Gene Tsudik and the anonymous referees of the USEC '07 workshop gave us valuable feedback on draft versions of the paper. Nitesh Saxena and Gene Tsudik helped initiate this work. Philip Ginzboorg and Kari Kostiainen helped in the design and implementation of the multi-device testing framework.

References

[1] J. M. McCune et al; “Seeing-Is-Believing: Using Camera Phones for Human-Verifiable Authentication”. In Proceedings of the 2005 IEEE Symposium on Security and Privacy, pages 110-124, 2005.

[2] N. Saxena et al; “Secure Device Pairing based on a Visual Channel (Short Paper)”. In Proceedings of the 2006 IEEE Symposium on Security and Privacy, pages 306-313, 2006.

[3] M. T. Goodrichet al; “Loud and Clear: Human-Verifiable Authentication Based on Audio”. In IEEE international Conference on Distributed Computing Systems, 2006.

[4] S. Vaudenay; “Secure Communications over Insecure Channels Based on Short Authenticated Strings”. In proceedings of the Advances in Cryptology, Lecture Notes in Computer Science volume 3621, pages 309-326, 2005.

[5] D. Balfanz et al; “Network-in-a-box: How to set up a secure wireless network in under a minute”. In USENIX Security Symposium, pages 207–222, 2004.

[6] D. Balfanz et al; “Talking to strangers: Authentication in ad-hoc wireless networks”. In Network and Distributed System Security Symposium, 2002.

[7] C. Kuo et al; “Designing an evaluation method for security user interfaces: lessons from studying secure wireless network configuration”. Interactions, 13(3):28-31, ACM Press, 2006.

[8] Near Field Communications forum, http://www.nfc-forum.org/home

[9] Wi-Fi Protected Setup Specification,Version 1.0, January 2007, available from https://www.wi-fi.org/published_specifications.php

[10] Bluetooth Special Interest Group, “Simple Pairing White Paper”, V10r00, August 2006 http://bluetooth.com/NR/rdonlyres/0A0B3F36-D15F-4470-85A6-F2CCFA26F70F/0/SimplePairing_WP_V10r00.pdf

[11] Wireless USB Specification Revision 1.0, Association Models Supplement Revision 1.0, March 2006, http://www.usb.org/developers/wusb/wusb_2006_0302.zip

[12] Microsoft, Windows Connect Now-NET Specification, September 2006 http://download.microsoft.com/download/a/f/7/af7777e5-7dcd-4800-8a0a-b18336565f5b/WCN-Netspec.doc

[13] C. Gehrmann et al; “Manual authentication for wireless devices”, RSA Cryptobytes, 7(1):29–37, Spring 2004. http://www.rsasecurity.com/rsalabs/cryptobytes/Spring_2004_Cryptobytes.pdf

[14] S. Laur et al; “Efficient Mutual Data Authentication Using Manually Authenticated Strings”, Cryptology ePrint Archive, Report 2005/424, 2005

[15] M. Čagalj et al; “Key Agreement in Peer-to-Peer Wireless Networks”, In Proceedings of the IEEE (Special Issue on Security and Cryptography), 92(2):467–478, February 2006.

[16] K. Kostiainen et al; “Framework for Comparative Usability Testing of Distributed Applications”, manuscript under preparation, 2007.

[17] G.A. Miller, “The magical number seven, plus or minus two: Some limits on our capacity for information processing”. Psychol. Rev. 63(2):81-97, March 1956.

[18] J. Palmer; “Attentional limits on the perception and memory of visual information”. Journal of Experimental Psychology: Human Perception and Performance, 16:332-350, 1990.

[19] R. Hammer et al; “Category Learning from Equivalence Constraints”. XXVII Conference of Cognitive Science Society (CogSci2005), July 2005.

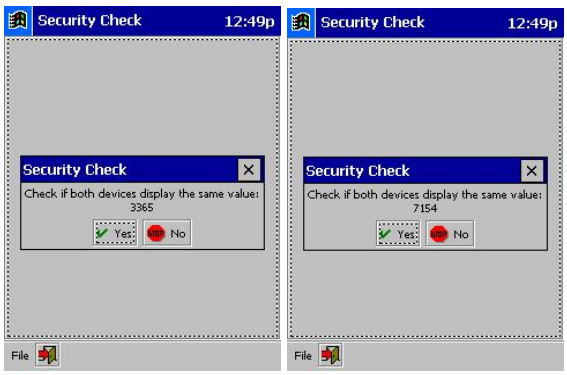
[20] D. A. Norman; “The Design of Everyday Things”. Doubleday, New York, NY, 1988.

[21] J. Suomalainen et al; “Security Associations in Personal Networks: A Comparative Analysis”, Nokia Research Center Technical Report NRC-TR-2007-004 available at http://research.nokia.com/tr/NRC-TR-2007-004.pdf

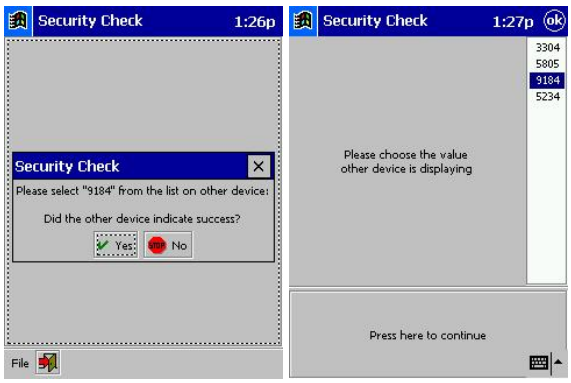
Appendix

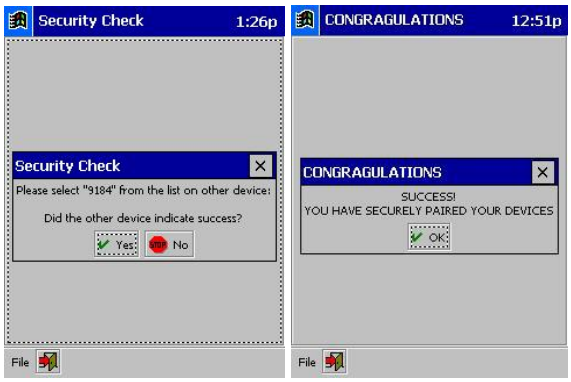
A. Screenshots From Round One Implementation

Compare-and-confirm:

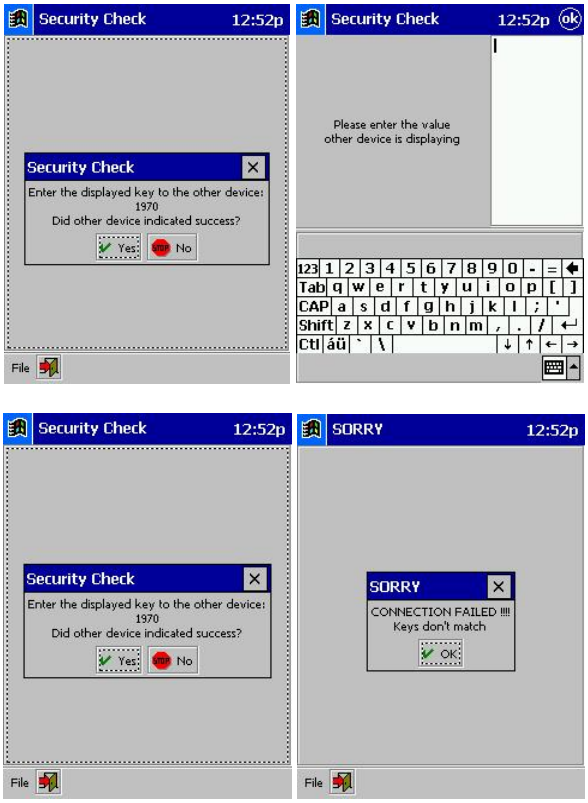


Select-and-Confirm (Selection and Confirmation Phases):

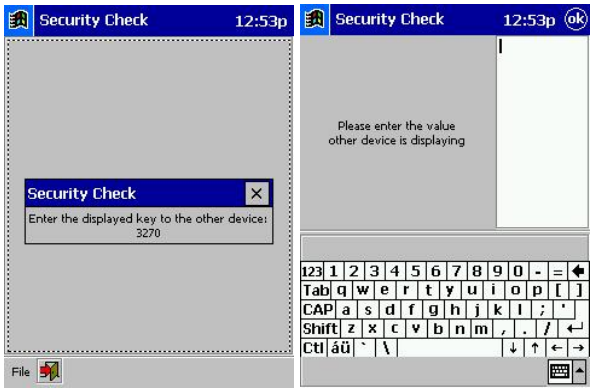




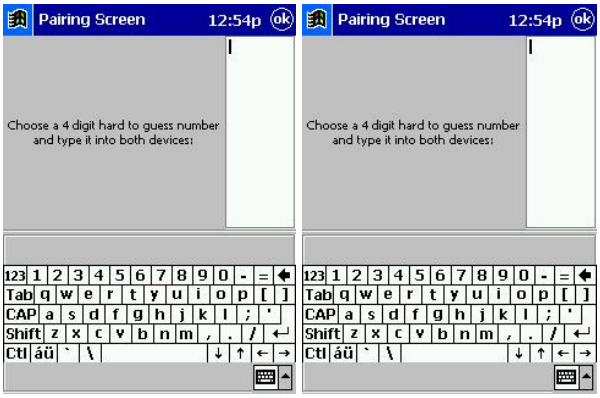
Copy-and-Confirm (Copy and Confirmation Phases):



Copy:



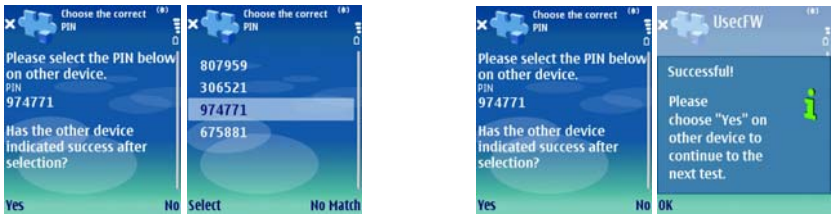
Choose-and-Enter:



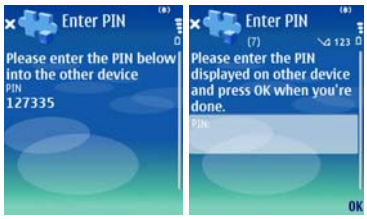
B. Screenshots From Round Two Implementation

Compare-and-Confirm

Select-and-Confirm (Selection and confirmation Phases)



Copy



Framework screen in between tests

C. Background Questionnaire

Demographics

Age

☐ 18-24 ☐ 25-29 ☐ 30-34 ☐ 35-39 ☐ 40+

Nationality

Native Language

Sex

☐ Male ☐ Female

Highest Grade Competed

☐ High School ☐ Bachelor ☐ Masters ☐ Doctorate

If you are college graduate, please list your major

Computer experience

For how long you have been using computers?

On a typical day, how many hours do you work with computers?

Mobile device experience

Do you have any personal mobile device such as cell phone, PDA, pocket pc, smart phone?

☐ YES ☐ NO

Is your mobile device capable of establishing any of Bluetooth, infrared or WI-FI connection?

☐ YES ☐ NO ☐ N/A

Do you use any of its Bluetooth, infrared or WI-FI functionality on a regular basis?

☐ YES (how often?) ☐ NO ☐ N/A

Please check the corresponding box if you have done it with your mobile device before:

☐ Playing two-player mobile phone games
☐ Using a wireless headset with your mobile phone
☐ Connecting your computer or PDA to the internet using your mobile phone
☐ Wirelessly synchronizing your mobile phone calendar with your computer calendar

I currently use wireless communication in my following devices

a.
b.
c.
d.

In general, I feel secure while using wireless communication

☐ Agree ☐ Disagree ☐ Neutral ☐ Don't Know

D. Post-test questionnaire

Please answer the following questions based on your experience using the methods. Where appropriate, we would appreciate if you would explain your answers and reasoning in the spaces provided or orally to us.

1. Please circle the method name that you think it is best described by the adjective

Easiest:	Compare	Select	Type
Hardest:	Compare	Select	Type
Professional:	Compare	Select	Type
Most secure:	Compare	Select	Type
Least secure:	Compare	Select	Type
Fun to use:	Compare	Select	Type
I would like my personal mobile device to be equipped with the following method(s):			
	Compare	Select	Type

2. I found the following aspects of certain methods very difficult to use

A.
B.
C.
D.
E.

3. I would prefer a method different from all of the above, or combination of those

☐ Yes (please explain it to us orally)
☐ No

4. I would find high level security useful in wireless connections to or from my mobile phone.

☐ Yes ☐ No

If yes, how would it be helpful / useful?

If no, why would it not be helpful / useful?

5. Please add any comments in the space provided that you feel will help us to evaluate the methods or come up with a better one. We would especially appreciate your input on comparing the methods from different perspectives. (You can answer this question orally if you would like to).

E. Orientation Script

Hi, my name is …….. I'll be working with you in this test session. Let me explain why we've asked you to come in today.

We're here to compare the usability of different implementations of a procedure that will take place on usage scenarios that needs communication between devices such as transferring photos or contacts. We'd like your help to understand the user point of view in this matter

You will be trying different protocols today which of all intend to the same thing with different interfaces. I'd like you to perform as you normally would why performing the requested operations. For Example, try to work at the same speed and with the same attention that you normally do. You may ask questions at any time, but I may not answer your certain questions if it may affect your preferences or we need to see how the protocols work with a person working independently.

During this session, I'll also be asking you to complete some forms and answer some questions. It is important that you answer truthfully. My only role here is to discover the usability flaws or advantages of each protocol from your perspective. So, don't hesitate to say your positive and negative opinions, critics or feelings. I need to know exactly what you think. Your careful evaluation of these mechanisms is very important and will help us build better products for tomorrow.

While we are working, I'll be standing/sitting nearby taking some notes. If you don't have any questions, let's begin by having you filling out the background and pre-test questionnaires.

F. More on Participant Profiles

Round One

- Do you have any personal mobile device such as cell phone, PDA, pocket pc, smart phone?
  - 100% YES
  - 0% NO
- Is your mobile device capable of establishing any of Bluetooth, infrared or WI-FI connection?
  - 45% YES
  - 55% NO
  - 0% N/A
- Do you use any of its Bluetooth, infrared or WI-FI functionality on a regular basis?
  - 27.5% YES
  - 17.5% NO
  - 55% N/A
- If your personal device is not featured with Bluetooth, infra-red or WI-FI, do you intend to buy such one within the next 6 months?
  - 20% YES
  - 35% NO
  - 45% N/A

Round Two

- Do you have any personal mobile device such as cell phone, PDA, pocket pc, smart phone?
  - 100% YES
  - 0% NO
- Is your mobile device capable of establishing any of Bluetooth, infrared or WI-FI connection?
  - 72.5% YES
  - 27.5% NO
  - 0% N/A
- Do you use any of its Bluetooth, infrared or WI-FI functionality on a regular basis?
  - 40% YES
  - 32.5% NO
  - 27.5% N/A
- I currently use wireless connection in my following devices
  - 75% PC
  - 40% Mobile Phone
  - 5% PDA
  - 7.5% Others

G. More on User Opinions

Round 1

- In general, I enjoy using high-tech products.
  - 50% Strongly Agree
  - 35% Agree
  - 15% Neutral
  - 0% Disagree
  - 0% Strongly Disagree

- In general, I feel comfortable and secure while using wireless communication
  - 7.5% Strongly Agree
  - 47.5% Agree
  - 30% Neutral
  - 15% Disagree
  - 0% Strongly Disagree
- Usability is a very important factor in my decision to use a certain new feature of a product.
  - 55% Strongly Agree
  - 37.5% Agree
  - 7.5% Neutral
  - 0% Disagree
  - 0% Strongly Disagree
- I found the following aspects of certain methods very difficult to use.
  - 80% Typing long numbers
  - 12.5% Coming up with a hard to guess number

Round 2

- In general, I feel secure while using wireless communication
  - 7.5% Agree
  - 47.5% Disagree
  - 30% Neutral
  - 30% Don't Know
- I would find high level security useful in wireless connections to or from my mobile phone
  - 60% YES
  - 40% NO (Main reasoning was not having any data that needs privacy)
- I found the following aspects of certain methods very difficult to use.
  - 42.5% Typing
  - 32.5% Comparing long numbers
  - 12.5% Except *Copy*, needing input in both devices
  - 12.5% Selecting from the list

- Correlation between adjectives in participants' view

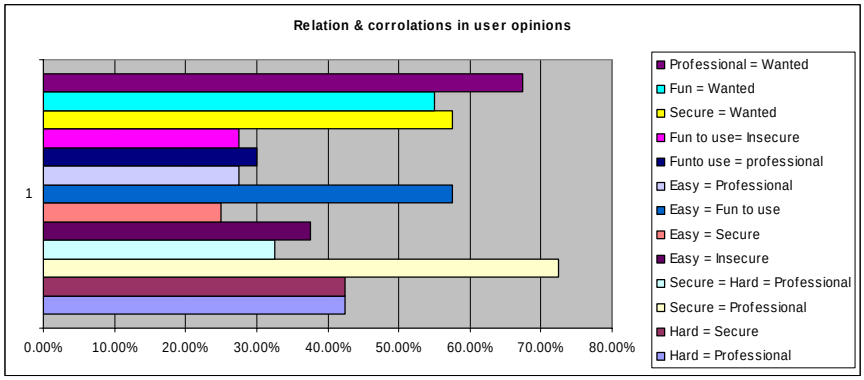


Fig. 3. Correlation analysis from test round two.