Two Heads are Better Than One: Security and Usability of Device Associations in Group Scenarios

Ronald Kainda
Oxford University Computing Laboratory
Parks Road, OX1 3QD, UK
ronald.kainda@comlab.ox.ac.uk

Ivan Flechais
Oxford University Computing Laboratory
Parks Road, OX1 3QD, UK
ivan.flechais@comlab.ox.ac.uk

A.W. Roscoe
Oxford University Computing Laboratory
Parks Road, OX1 3QD, UK
bill.roscoe@comlab.ox.ac.uk

ABSTRACT

We analyse and evaluate the usability and security of the process of bootstrapping security among devices in group scenarios. While a lot of work has been done in single user scenarios, we are not aware of any that focusses on group situations. Unlike in single user scenarios, bootstrapping security in a group requires coordination, attention, and cooperation of all group members. In this paper, we provide an analysis of the security and usability of bootstrapping security in group scenarios and present the results of a usability study on these scenarios. We also highlight crucial factors necessary for designing for secure group interactions.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems—*Human Factors*

General Terms

Experimentation, Security, Human Factors

Keywords

Security Protocols, Usability, Device Association, Group Interactions

1. INTRODUCTION

One of the challenges in computer security is the sharing of cryptographic keys among legitimate participants. Public key cryptography has made sharing of cryptographic keys relatively easy to achieve, however public keys still need to be authenticated before use. The use of trusted third parties or a Public Key Infrastructure (PKI) in authenticating public keys is currently standard practice in many applications including e-commerce.

Symposium on Usable Privacy and Security (SOUPS) 2010, July 14–16, 2010, Redmond, WA USA

In *ad hoc* wireless networks of mobile devices, no PKI or trusted third party is practical or sufficiently universal to cover secure sharing of cryptographic keys [26]. This has led to research into new ways in which devices in *ad hoc* wireless networks can establish secure communications without relying on trusted third parties or a PKI. Among the results of this research is the proposal for using two channels: a high bandwidth (normal) channel, which is subject to the Dolev-Yao attack model [6] and a low bandwidth Out-Of-Band (OOB) channel. Messages on the normal channel may be modified, deleted, or spoofed by an attacker. On the other hand, messages on the OOB channel cannot be modified, deleted or spoofed.

In this proposal, the associating devices exchange public information, such as public keys, using the normal channel. Either a single device, whose key needs to be authenticated, sends its public key to an authenticating device(s), or all devices involved exchange their public keys. Devices then independently compute a cryptographic fingerprint of the information exchanged. The fingerprints are compared via an unspoofable OOB channel to verify the authenticity of the public key(s). A requirement on the OOB channel is that integrity of information is maintained but not secrecy [24].

Bootstrapping security in *ad hoc* networks for groups differs in many respects with pairwise scenarios. The increased number of devices also increases the chance of these devices being significantly different in terms of affordances, computation ability, and other features. Parallel to these differences among devices are the differences among humans using the devices.

There is a gap in the research to understand the challenges of group association scenarios. Studies on pairwise associations have identified challenges and proposed improvements specific to these scenarios. Given the differences between pairwise and group association scenarios, the challenges identified, and recommendations made for pairwise association scenarios may not be applicable to groups. It is, therefore, timely that this work attempts to highlight some of the issues in device associations for group scenarios.

The paper is organised as follows; related work is presented in Section 2 with a discussion of group scenarios in Section 3. Experiment design is presented in Section 4 with results presented in Section 5. We discuss results in Section 6 and conclude in Section 7.

2. RELATED WORK

2.1 OOB channels

Various OOB channels for transferring or comparing short strings have been proposed. The channels may be grouped into five main categories: manual comparison, manual copying and entering, using auxiliary devices, short range directed channels, and timing methods.

2.1.1 Compare and confirm

Compare and confirm involves users comparing strings displayed on two or more devices and indicating on devices involved whether compared values match or not. The string may be encoded into one of the various forms for easy comparison by users. Proposed forms include images [25], sentences [10], numeric [11], and sounds [33].

Studies on usability of device association methods in pairwise scenarios have found that *compare and confirm* is the most usable and most preferred method [17, 32]. The method, however, is susceptible to security failures as users are not, in any way, obliged to make an effort and compare fingerprints accurately.

2.1.2 Copy and enter

This method involves having one device display a fingerprint which a user copies to another device(s). The device where the fingerprint is entered then compares the entered value with its own. After comparison, the device then indicates whether the values match or not. Proposed forms for representing fingerprints include numeric and alphanumeric. *Copy and enter* is commonly used in Bluetooth device pairing [11].

Copy and enter is not susceptible to security failures as the comparison of fingerprints is carried out by devices rather than humans. Compared to *compare and confirm*, it demands much more effort from users resulting in participants in usability studies (e.g. [17, 32]) ranking it below *compare and confirm*.

2.1.3 Auxiliary device

Auxiliary device methods depend on other devices to transfer or compare limited information between devices. A device may be used as a transfer medium only or as a verifier too. For example, devices with displays may encode fingerprints of information they want to be authenticated into a 2D-barcode and an extra device (not participating in the association), with a digital camera and appropriate software, may read the barcodes on all devices involved and compares them. The user indicates on the associating devices the response from the auxiliary device. In some cases, one or more devices participating in the association may have the capability to act as an auxiliary device and thereby avoid the necessity of an extra device. Proposed devices include data cable [34], a digital camera [23], and external storage media such as memory cards.

2.1.4 Short range directed channels

These methods depend on short range directed data transmission technologies to transfer information between devices. They require users to only align associating devices to facilitate exchange of information between devices. Proposed methods include infra-red [2, 8], light [20, 22, 28], and integrity regions [3].

In general, short range directed channels are criticised for taking humans out of the loop. Users are unable to verify the authenticity of a transmission. These methods can also only be applied in cases where associating devices are close together, say a few centimetres apart.

2.1.5 Timing

Timing methods rely on transmission of cryptographic information within well timed intervals such that an intruder finds it hard to synchronise and successfully attack an association. Proposed methods include shaking devices [21] and pressing buttons [29].

These methods are only feasible in cases where users are capable of carrying out required tasks. For example, the device shaking proposal is not feasible if one of the devices involved is a vending machine. In group scenarios, these methods may be challenging, considering users' actions must be synchronised in order to have a successful association.

2.2 Usability studies

There have been efforts to evaluate usability of OOB channels. Uzun *et al.* [32] conducted a usability study on 4 different types of OOB channels; *compare and confirm, copy and enter, select and confirm,* and *choose and enter*. They found that *compare and confirm* was rated as the easiest method with *copy* ranking last while ranked as the most secure and personal choice.

More recently, Kobsa *et al.* [17] conducted a study on a number of OOB channels. They recommended different methods for different device affordances. Their study, however, was only concerned with usability of methods tested and, therefore, missed an opportunity to evaluate the security of those methods.

In our earlier work [13], we conducted a study on nine different methods for device association. In this study, *compare and confirm* was ranked highest. However, *copy and enter* was the recommended method based on the fact that it is not subject to security failures while *compare and confirm* is.

All the above studies focus on single user pairwise associations. No study to date has focussed on multi-user scenarios to understand the impact of these scenarios on usability and security of OOB channels.

3. DEVICE ASSOCIATION IN GROUPS

3.1 Association scenarios

Secure device association for groups may be characterised into four scenarios.

1. **One-to-many**
   One of the scenarios of group associations is one-to-many. In this scenario, one device is authenticated by two or more devices. The authenticated device may be manned such as another mobile device or unmanned such as vending machines. In either case, users authenticating the device will need an OOB channel from the device to their devices.

   A practical example of this scenario is a medical emergency. In a medical emergency, say an earthquake with several victims, first responders will attend to survivors before taking them to a nearest available

medical facility. In order to provide efficient and effective service at the hospital, first responders may need to transmit information to the medical facility so that medical stuff at the hospital are prepared for coming victims. However, first responders may not have devices powerful enough for long range transmission of data and may want to create a local area network among their devices with only one powerful device through which other devices transmit information to the medical facility. The crucial factor is for first responders to ensure that their devices are connected to the right transmitter—in which they have to authenticate it before any information is sent. In this life and death situation, the authentication process must be efficient.

Another example scenario is a game, poker for example, in which players play the game using mobile devices while it is centrally managed by a single device. Individuals who want to play together form a group and authenticate the control device to ensure their devices are connected to the correct control device. The control device could be handling several groups, hence a number of participants participating in a single session of device association form a group that is managed independent of other groups.

2. **Many-to-one**
   The second group association scenario is many-to-one. In this scenario, one device is authenticating several. This may be appropriate in scenarios where group membership is controlled by one individual.

   One example of such a scenario is a meeting where a group controller wants to share sensitive information with other participants. For example, a CFO wanting to share sensitive information with members of a marketing team would want to have control on the attendees and would also endeavour to ensure that the information is only shared with known participants and none else.

3. **Partial symmetric**
   Partial symmetric is a congruence of many-to-one and one-to-many. In this scenario, a group of device authenticate one device and vice versa. Group members authenticate a single device by ensuring that fingerprints displayed on the devices match with one displayed the single device. The single device owner authenticates group members by checking that devices of group members display *success* on their devices.

   To illustrate partial symmetric, consider a vending machine that issues cinema tickets. The machine can issue multiple tickets at a single instance to facilitate for group orders. Each member of a group will receive a digital ticket on their device because the gate to the cinema allows for a single entrant hence one person cannot receive tickets on behalf of other.

   In this example, group members will authenticate the vending machine to ensure that they do not receive fake tickets from a rogue device. On the other hand, the vending machine does not immediately issue tickets until a person manning the machine or one of the group members indicates on the vending machine that the association was successful. If the vending machine does not wait for instruction to distribute tickets, it may send them to people who are not members of a group ordering tickets.

4. **Full symmetric**
   The previous two scenarios require authentication in one direction or from a group of devices to one and vice versa. There are scenarios, however, where each device participating in the association needs to authenticate every other device. This may be viewed as a many-to-one scenario repeated for each device in the group. Each participant in the group authenticates other members by ensuring that their fingerprint matches every other device's. An example for full symmetric is a multi-player game where there is no central device to which participants' devices can connect. In this scenario, each participant is keen to ensure that the game is played with only the people he is seeing and no one else.

3.2 Security and Usability challenges of group scenarios

The number of users involved in bootstrapping security between mobile devices may have serious implications on the security and usability of OOB channels. We categorise the number of users as either single-user (where one person controls all devices involved) or multi-user (where each device has its own user). In a multi-user scenario, a well designed OOB channel may consider distributing the amount of work among participating users and, as such, it may give an opportunity for using fingerprints of sufficient size (for theoretical security) as opposed to where a single user is expected to do all the work.

As security is a process rather than a product [31], the number of nodes where security may fail increases with each additional device or device/user pair since the correct behaviour of all participants is necessary to achieve desired security [5]. In secure device associations, participants achieve global security—by sharing a common cryptographic key, for example—among them only when they all behave correctly and are diligent in detecting anomalies. OOB channels, therefore, can only achieve security when they make desired user actions easier to do than undesirable ones within the context in which they are used as they are part of a larger system.

The concerns to be addressed here are: how can we design (or how do we propose) OOB channels that allow for distribution of human effort among participants? How can a single user establish a secure association of multiple devices with acceptable mental and physical effort? How does increasing the number of devices or device/user pair affect the usability and security of a particular OOB channel?

We categorise authentication as either one-way (asymmetric) or mutual (symmetric). In one-way authentication, one device authenticates one or more participating devices. For example, an Access Point (AP) authenticating mobile devices wanting to access the Internet through it (assuming the AP is configured to authenticate devices). In this scenario, a user may be happy to identify the AP by name (if they know it) or by other means. In short, the user conducts a weaker authentication of the AP. The AP on the other hand requires a stronger authentication in which it may require the user to transfer some information, using an

OOB channel, to verify that the owner of the device is within the vicinity and hence (presumably) has access rights to it.

In mutual authentication, however, each of the participating devices authenticates all the other devices. In the AP example, the user or their personal device may require more than just a name of the AP. The device may require the AP to compute something which the user can verify.

Either of these scenarios poses different usability challenges. In one-way authentication, the authenticating device's acceptance of an association request is good enough for the authenticated device. For example, once a connection to a named AP is established, that is good enough for the device. In practice, the AP may require the user to transfer some information from the AP to the device and no further action from the user.

In mutual authentication, a user(s) may be required to take extra steps. The AP may be required to indicate to the user acceptance or refusal of the association request and indicate to the device appropriately. The amount of effort expended in mutual authentication may be double that expended in one-way authentication. For example, using a 2D barcode (as in [23]) to encode the fingerprint of exchanged information, the barcode may have to be captured *n-1* times for one-way authentication and *n(n-1)* times for mutual authentication where n is the number of devices participating in the association. Understanding this difference in human effort between the two scenarios may be invaluable to designing usable OOB channels.

The extra step in mutual authentication is not only an increase in human effort but also a step where security may fail. For example, a user misinterpreting a refusal by the AP as an acceptance of the association may result in pairing the device to an unintended AP or the user may interpret the message on the AP correctly as a refusal but fail to indicate accordingly on the device.

Another challenge to security and usability of bootstrapping security in group scenarios concerns group size. A hidden node may participate in the bootstrapping process exposing all shared secrets. To prevent this attack, participants need to ensure that the number of devices participating in the association is what is expected. This may be a job of initiator, for groups with a leader, or every group member ensures that only expected number of devices are communicating with his/her device. It is a usability challenge because it will require participants to verify the number of devices involved. In applications where the number of devices can be predetermined, it may be reasonable to set this application level without needing users to verify.

In device association of groups, two channels of communication are important. First, an initiator (personal controlling the group) must be able to communicate to each group member. This communication may carry information about fingerprints, status of association or of other group members. The initiator may benefit from a broadcast channel where one transmission gets to all group members. For example, to announce the value of the fingerprint initiator may read it loudly for everyone else in the room rather than passing their device to each participant to check the fingerprint.

The second communication channel important in device associations for groups is from each group member to initiator. Group members need to communicate the result of the security task performed, either it succeeded or not, and any relevant communication that may help in the association process. Without this channel, initiator will be in no position to know the status of the association once she reads out a fingerprint for other to compare and enter.

With this background, some of the sources of security and usability problems in group scenarios are as follows:

- **Failure of communication from initiator to group members:** when initiator fails to communicator correctly, group members may take wrong information which may result in devices rejecting legitimate associations. This may cause frustration as the process has to restart after a failure.

- **Failure of communication from group members to initiator:** group members must communicate results of an association to initiator for the latter to make the correct decision of either accepting or rejecting an association. A failure in communication may result in usability problems because initiator may reject perfectly valid associations and also in security problems because initiator may accept invalid associations.

- **Inattentiveness by initiator:** Initiator should be attentive and interpret messages from group members correctly. Failure to do so will result in similar problems as discussed in the previous bullet point.

- **Inattentiveness by group members:** this is similar to the problem in the first bullet point except that in this instance, group members do not pay attention to initiator's messages.

4. EXPERIMENTAL DESIGN

Our study used the partial asymmetric association scenario for a number of reasons:

1. It covers both one-to-many and many-to-one scenario. We can, therefore, using a single study evaluate performance of OOB channels for both scenarios covered in partial symmetric.

2. Choosing only one-to-many or many-to-one limits the generalisation that we can draw from data. For example, data on one-to-many association scenario may not be extended to any other scenarios.

3. Full symmetric is a special form of, and can be achieved using, partial symmetric. For example, in partial symmetric rather than having participants report the status of the association to initiator, they may report it to other group members as well.

In order to evaluate security and usability of OOB channels in group scenarios, we identified possible sources of both usability and security problems based on [16]. For usability, the following were identified; effectiveness, efficiency, satisfaction, and accuracy. These elements are commonly used as metrics in usability studies. For security we identified the following; attention to the association process, conditioning, social context (group), vigilance (can participants be actively attentive to the association process throughout), and motivation. In the design and conducting of the study, particular attention was paid to these elements.

Upon identification of elements that may pose challenges to security or usability or both, we followed the process for

Figure 1: Group interactions in device association

evaluating usability and security of secure systems proposed in [16]. This process was used because, rather than just paying attention to usability in the design and conducting of the study, it allowed us to pay attention to security issues too. Using this process, we identified usage scenarios—scenarios that represent real world applications of a secure device association rather just security tasks as these are a secondary goal to users. We used the following usage scenarios; exchanging contacts, digital cash transfer, group messaging, and group quiz. We then identified threat scenarios; things that we do not want to happen in a secure system. In secure device association in a group scenario, threat scenario are; accepting a non-matching fingerprint, initiator interpreting failure of device association from one or more devices as success, intruder joining network without knowledge of initiator or other group members. We designed these threat scenarios into our study so as to determine how likely users may detect and defeat them.

4.1 Participants

To increase power and reduce variability, we used a repeated measure with counterbalancing (to minimise learning effects). We recruited 49 participants (24 male, 25 female). Participants were randomly grouped into 13 groups (See Figure 1) with group sizes of 2 (2 groups), 3 (4 groups), 4 (3 groups), 5 (3 groups) and 6 (1 group). Each group performed the same test conditions (in different orders) and primary tasks. In each group, one member was randomly assigned to be initiator. One group of 2 participants was later excluded due to errors in data collected. Table 1 summarises 47 participants (excluding 2 males as above) demographics.

Gender	Male: 46.7% Female: 53.3%
Age	18 - 25 51.1% 26 - 35 21.3% 36 - 45 17% 46 - 55 8.5% 56+ 2.1%
Education	High School: 19.1% College: 31.9% Graduate: 27.7% Postgraduate: 21.3%

Table 1: Participant demographics

4.2 Materials and apparatus

In conducting the study, we implemented a tool that incorporated logging of user actions and user interfaces for interacting with mobile devices. The tool was implemented using Java Micro Edition (J2ME) and runs on mobile devices that support Mobile Information Device Profile (MIDP) framework implementations. It supports test configurations (such as number of tests to run), event logging (i.e. completion time, number of buttons pressed, security and non security failures), different experimental designs (e.g. randomised, counterbalanced), and error simulation. We implemented all usage and threat scenarios into the tool. We ran the tool on Nokia N95 and Blackberry Bold 9000 devices.

After Scenario (ASQ) [19] were used to capture user ratings for each method immediately after encountering the method. For each method, ASQ capture data on three main components of usability (satisfaction, efficiency, and effectiveness):

- satisfaction with the ease with which a method was used,
- satisfaction with the amount of time spent on a method,
- whether a participant felt they could effectively carry out primary tasks using a particular method.

ASQ is a rating scale type questionnaire consisting of 3 questions with answers based on a scale of 1 to 7, with 1 corresponding to *strongly agree* and 7 to *strongly disagree*. Many rating scales use a scale of 5 intervals rather than 7. However, it has been found that reliability of rating scales increases with the number of items and also the number of interval points for each item, and levels off at about 7 intervals with no significant increase after 11 intervals [19], hence the use of a 7 point interval scale.

An End of Experiment (EoE) questionnaire gave participants an opportunity to identify methods they felt were easy, difficult and which ones they preferred or would avoid. It also asked participants to rank each method on a 7 point scale with 1 corresponding to *very difficult* and 7 being *very easy*. Interviews gathered participants' views and comments on what they felt about the methods and group interactions.

Each test session lasted for about an hour, including a discussion. The sessions were video taped so as to analyse and understand how participants interacted and help identify elements that may help or hinder secure device association in these scenarios.

4.3 Methods tested

We chose methods to be tested based on previous studies of pairwise device associations (e.g. [13, 17, 32] and on their feasibility to group scenarios. We also assumed that in group scenarios, devices will have reasonable input/output interfaces. Based on this, we tested the following methods - *compare and confirm - numeric, compare and confirm - images, copy and enter - numeric, repeated comparison - numeric*, and *word-matching and number-typing.*

4.3.1 Compare and confirm - numeric

Among the proposed forms of presenting fingerprints to users in *compare and confirm*, numeric is the most basic and participants in previous studies (such as [32] rated the method as the most usable. We included this method in order to assess whether users may still find it usable in group

settings. In addition, we were interested in assessing the security of the method considering that it is subject to security failures.

Cases tested:

1. Normal case: every device in a group displayed a fingerprint that matched one shown on initiator's device. Comparing correctly and selecting correct commands, association was to be successful. A mistake by a participant may result in initiator rejecting an association, that is, safe failure.

2. Failure case: one device displayed a fingerprint that did not match with initiator's and an accurate comparison should result in initiator rejecting an association similar to the previous case. If a participant does not pay attention and does not communicate the failure such that initiator accepts the association, it results in a security failure.

4.3.2 *Compare and confirm - images*

In our previous study, participants had difficulties comparing images on two mobile phones, especially when the same image was displayed on both devices. It was realised that this was the case because participants where looking for differences between the two images. It was later on suspected that this could have been due to instructions given to participants—asking "ARE THE IMAGES DIFFERENT?". For group scenarios, it was decided to change instructions to "ARE THE IMAGES SAME?" to see whether this would improve the performance of image comparison. During preliminary studies it was, however, decided that this method be dropped and not be used during the actual study because changes in instructions did not change participant's behaviour towards the method and it was taking much longer compared to other methods.

4.3.3 *Copy and enter - numeric*

This method is not susceptible to security failures and we were interested in assessing its usability in a group context. Though there has been proposals to use other formats, such as alphanumeric, only numeric was used because it is easier to type on a multi press keypad such as one found on standard mobile phones compared to alphanumeric.

Cases tested:

1. Normal case: every device in a group had the same fingerprint as initiator and upon typing it in correctly, an association should be successful. An incorrect entry will result in initiator rejecting the association causing a safe failure.

2. Failure case: to simulate a failure case for this method, one device is randomly selected and rejects any string of numbers entered. For example, a participant may enter a number correctly as read by initiator but the device will alert user that the association failed. While in practice a successful attack may be impossible to carry out for this method, the aim of the failure case was to assess whether users can respond correctly in such an event. In addition, failure to communicate to initiator that association failed may result in initiator accepting an association with a wrong device(s), that is security failure.

4.3.4 *Repeated comparison*

To compel users to carry out comparison without undue effort, *repeated comparison* was proposed in [14] as a two step comparison process. In addition to a fingerprint, an authenticating device generates a random string of similar format as the fingerprint. The authenticating device then randomly chooses to display either its fingerprint or the random value. The user compares and indicates whether the string displayed on the authenticating device matches that on the other device. An authenticating device then displays the remaining string and the user does the comparison again. An authenticating device accepts a connection only when a user indicates a match for a fingerprint and a mismatch for the random value. The argument for this method is that, unlike compare and confirm, it is difficult for users to ignore the crucial task of comparing fingerprints without causing one device to refuse connection.

Cases tested:

1. Normal case: every member device in a group had an actual fingerprint and a correct comparison by participants would result in a successful association.

2. Failure case: one device in a group was randomly selected by initiator device and assigned values of which none is a match to an actual fingerprint. A participant with this device saw (after correct comparison) a "connection failed" message which should be communicated to initiator. Failure to communicate this message, initiator may accept association—when in fact one device in the group has rejected it—resulting in a security failure.

4.3.5 *Word-matching and number-typing*

This method is based on the fact that *copy and enter* is not subject to security failures but is regarded as difficult to use. While earlier work has argued that typing short strings on devices with limited input interfaces is hard for most users, the popularity of Short Message Service (SMS) is an indication that users are comfortable with such a task. We are, however, cognisant of the lack of motivation from users to type strings for the sake of security. *Word-matching and number-typing* is, therefore, aimed at offering the same level of security as *copy and enter* but only requiring users to type a smaller number of digits.

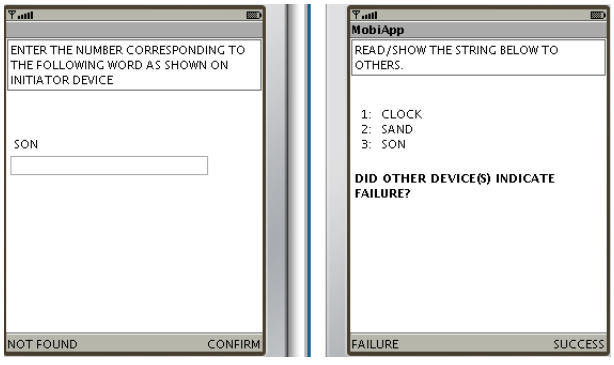

Figure 2: Word-matching and number-typing method: group member device on left and initiator's on right

Figure 2 shows *word-matching and number-typing*. Initiator's device displays 3 words, two of which represent an actual fingerprint. Group members' devices randomly display one word from a computed fingerprint and asks users to enter the position of the word shown as displayed on initiator's device. In Figure 2, for example, a user will type '3' for SON and press confirm. The device will display a word and a user enters the position of that word as well.

Tested cases:

1. Normal case: all devices display two words, one after another, that are among a list of 3 displayed on initiator's device. Correct entry of positions of these words on member devices results in successful association. Mistakes in entry results in a safe failure.

2. Failure case: one randomly selected device displays a word that is not among words displayed on initiator's device. This results in a security failure if it is not communicated to initiator so that the association is rejected.

4.4 Participant tasks

As earlier stated, we developed usage scenarios (representative tasks) to represent real world applications in which our studied security tasks may be applied. Upon arrival in our lab, participants were taken to a room where the study was conducted. They sat around a square table and were asked to sign a consent and enrolment forms. Mobile devices were then distributed randomly to participants, except Blackberry devices that were only given to participants who had used one before. Participants were then given an overview of what the study was about and what the tasks were, outlining the roles of initiator and other group members.

During the study, participants were allowed to ask the test observer or discuss amongst themselves. Mobile devices asked participants to complete ASQ questionnaires as they encountered each method. This was deliberately done so that these questionnaires were completed while users still had a vivid picture of a method a particular questionnaire was about.

After a successful connection among devices, initiator then started an application (primary task). Upon completion of the primary task, another connection process was initiated. Figure 3 shows screen shots of participants' tasks. The tests were designed such that there were two normal cases and one failure case for each method. We used two normal cases for each method so as to check if there will be an improvement the second time a method is encountered. In addition, 3 of the 8 normal were meant to result in a failure due to a wrong number of devices being displayed on initiator's device.

4.5 Hypotheses tested

Three hypotheses were formulated:

- There is no difference between different age groups in terms of completion times — previous studies (e.g. [17]) have found that older participants take longer time to complete device association compared to younger ones.

- There is no difference between different methods in terms of completion times.

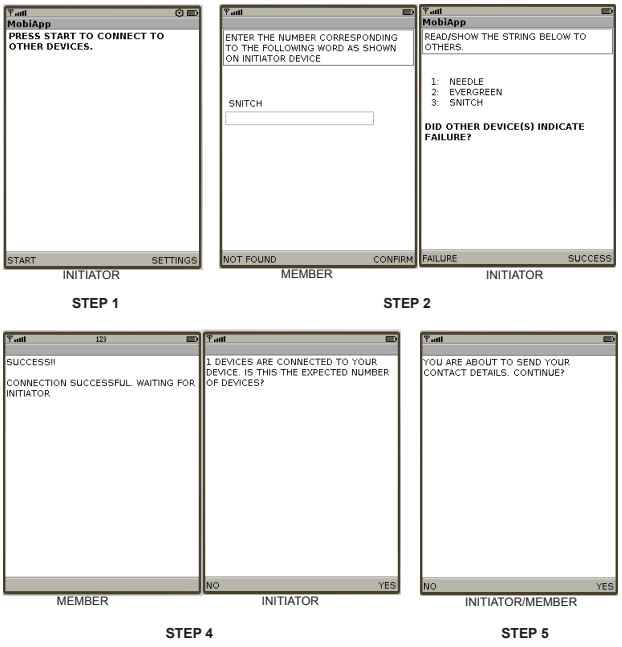

Figure 3: Task senquences: Initiator starts a connection and carries out a security task (in this example *word-matching and number-typing*) with group members. After successful association, initiator confirms number of devices and activates a primary task, in this case exchanging contacts

- There is no difference between different methods in terms of rating scores.

5. RESULTS

As hypotheses above state, we were interested in learning whether there are differences in performance between different age group and whether there are differences in performance and rating scores between different methods. Results are, therefore, analysed first by age and then method.

5.1 Analysis by age

We split participants into two age categories: 35 years old and below (n=34) and over 36 (n=13). Of interest are the times to complete security tasks, errors committed, ASQ scores (ratings), and preferences. To calculate the average completion times, initiator times were eliminated. This is because initiator completion times were significantly higher than other group members and there were more initiators in the younger group compared to the older group. Errors were calculated as number of errors committed divided by maximum possible and converted to percentage. Modal scores are used for ASQ scores (ordinal data) while preferences were calculated as percentage of age group who preferred a particular method. The results are summarised in Table 2.

We used a t-test to evaluate the statistical significance in completion times between the two groups for each of the methods. The test showed no significant difference in completion times for *compare and confirm* with p (2-tailed)= .666, *repeated comparison* with p (2-tailed) = .185, and *copy and enter* with p (2-tailed) = .414. There was, however, a significant difference for *word-matching and number-typing*

	ASQ (Mode)		Failures (%)		Time (Seconds)	
	Y	O	Y	O	Y	O
Compare and confirm	7(8)	6(5)	0	0	7	8
Repeated comparison	6(9)	6(3)	9	14	14	19
Copy and enter	7(9)	7(5)	5	0	12	13
Word-matching and number-typing	6(5)	6(3)	3	11	18	26

Table 2: Performance by age: Y = younger group (<36 years, n=27), O = older group (>35 years, n=9). X(Y): X = mode, Y = frequency

with p (2-tailed) = .024.

5.2 Analysis by method

We summarise the performance of each method according to security and non-security failures, completion times, participants' rating scores, and preferences.

5.2.1 Security and non-security failures

Non-security failures are events where device association is terminated by initiator either because one member made a mistake on the security task or because of miscommunication between initiator and members. Security failures result in device association when it is supposed to fail. Security failures may occur at two levels; when one device indicates unsuccessful association but initiator indicates success on their device or when device association is successful but number of devices connected to initiator device is not what is expected and initiator fails to notice. Table 3 summarises the results on security and non-security failures.

	Security %	Non-security %
Compare and confirm	0	0
Repeated comparison	2.4	17.9
Copy and enter	0	3.6
Word-matching and number-typing	0	8.3

Table 3: Security and non-security failures

None of the security failures observed were due to accepting wrong number of devices connecting to initiator but rather due to failure in communication between group members and initiator. Failure in communication was also partly the problem with non-security failures. With *repeated comparison*, the high percentage of non-security failures was due to group members misunderstanding the method; they would compare the first number displayed on the device and rather than comparing the second too, they reported this directly to initiator. For example, one participants reported, *"oh, I had a number which is same as yours [initiator] but now I have a different one"*, with initiator responding *"OK, that is a failure then"*. *Copy and enter*'s failures were due to typos while there were some confusions with *word-matching and number-typing* which caused some participant's to type same digit for both words.

5.2.2 Completion times

We analyse completion times at two levels; a group member's time to compare or type fingerprints and initiator's time to confirm success or otherwise. Initiator's completion time represents a group's time to complete an association—the time from when a fingerprint is displayed on their device to when every member has completed the security task and communicated the result to initiator. Tables 4 and 5 summarise the results for group members and initiators respectively.

	Min	Max	mean
Compare and confirm	2	42	7.89
Repeated comparison	3	64	17.63
Copy and enter	5	46	12.97
Word-matching and number-typing	6	94	22.89

Table 4: Group members' completion times

A repeated measure analysis of variance (ANOVA) was used since this is a within-subject design with more than two dependant variables. Mauchly's test indicated that the assumption of sphericity had been violated, χ^2 (5) = 74.36, p < .05, therefore a corrected value (Greenhouse-Geisser correction) of F was used. The test showed that there are significant differences in completion times among methods F(2.084,216.77) = 36.6 and p = .000. Pairwise comparisons also showed that each methods completion times were significantly different from each of the other methods with p-values ranging from .000 to .017.

	Min	Max	mean
Compare and confirm	7	278	40.97
Repeated comparison	11	105	33.94
Copy and enter	8	107	36.27
Word-matching and number-typing	11	147	48.27

Table 5: Initiators' completion times

A repeated measure ANOVA test on completion times for initiators showed that there are no significant differences between methods with F(2.04,71.3) = 1.22 and p = .277 (Mauchly's test indicated that the assumption of sphericity had been violated, χ^2 (5) = 31.24, p < .05, therefore a corrected value, Greenhouse-Geisser correction, of F was used). This seems contradictory with earlier analysis on group members where differences in completion rates are significant. Looking at video evidence, however, shows that initiators allowed some time to elapse before asking group members if their devices displayed failure or success. In some cases, this time was after members have completed their tasks, while in others they were still undertaking the tasks. To some extent, no matter how fast group members completed their tasks initiators allowed for some time before they thought it was time to move to the next one hence the lack of significance between methods.

5.2.3 Rating scores

We analyse initiator and other group member ratings separately. This is because of the difference in tasks carried by each group on each method. For example, while group members are required to type 6 digit numbers on their devices in *copy and enter*, initiator only reads the number displayed on their device. Tables 6 and 7 summarise the results for

group members and initiators respectively. The tables show the minimum and maximum scores (and their frequencies) for both ASQ score and overall (O) rating scores.

For group members, the table indicates that participants changed their rating scores between initial encounter with a method and completion of study. For example for *compare and confirm*, initially only 10 participants rated the method with a score of 7 and 19 gave the same score on the overall scale.

	Min	Max	Min(O)	Max(O)
Compare and confirm	5(3)	7(10)	4(1)	7(19)
Repeated comparison	4(2)	7(6)	2(1)	7(13)
Copy and enter	4.3(1)	7(13)	4(4)	7(19)
Word-matching and number-typing	4.3(1)	7(6)	4(2)	7(18)

Table 6: Group members' rating scores. X(Y): X = score, Y = frequency. Min(O) = min for overall score

We analysed rating scores for statistical significance using Friedman test since the data is ordinal. The test showed that there are significant differences in rating scores (ASQ) between methods with $\chi^2(3) = 11.655$ and $p = .009$. The test ranked *copy and enter* first, followed by *compare and confirm*, *word-matching and number-typing*, and finally *repeated comparison*. A pairwise Friedman's test was also carried out to find which methods had significant differences between them. The tests showed that there is significance between *copy and enter* and *repeated comparison* with $\chi^2(1) = 8.91$, $p = .003$ and between *copy and enter* and *word-matching and number-typing* with $\chi^2(1) = 4.84$, $p = .028$.

A Friedman test on overall scores, however, shows no significance between methods with $\chi^2(3) = 5.526$ and $p = .137$. Again, this statistic just shows that participants changed their ratings—by the end of the study they had a better understanding and grasped the tasks required of them hence more participants gave favourable scores thereby normalising initial differences.

We expected initiators to give more high ratings compared to group members considering the difference in the tasks they performed. We also expected similar tasks, e.g. reading a number, to be similarly rated. Results, however, show that this was not the case. First, a Friedman test shows that there is no significant difference in ASQ scores ($\chi^2(3)=4.558$, $p = .207$) while there is significance in overall scores ($\chi^2(3)=11.082$, $p = .011$). For both scores, *copy and enter* was ranked first, followed by *repeated comparison*, *word-matching and number-typing*, and lastly *compare and confirm*.

	Min	Max	Min(O)	Max(O)
Compare and confirm	3.7(1)	7(3)	3(1)	7(2)
Repeated comparison	4.3(2)	7(4)	2(1)	7(5)
Copy and enter	2(1)	7(4)	5(1)	7(8)
Word-matching and number-typing	3(1)	7(4)	3(1)	7(4)

Table 7: Initiators' rating scores. X(Y): X = score, Y = frequency. Min(O) = min for overall score

A pairwise Friedman test shows that there is significance between *copy and enter* and *repeated comparison* with $\chi^2(1)=4$, $p = .046$, *copy and enter* and *word-matching and number-typing* with $\chi^2(1)=5$, $p=.025$, and between *copy and enter* and *compare and confirm* with $\chi^2(1) = 6$, $p = .014$.

5.2.4 Preferences

As in rating scores, we analyse preferences in terms of initiator or group member. Participants were asked to indicate all methods that they felt were easy as well as those they felt were difficult. They were also asked to indicate which method they felt was easiest, most difficult and which method is their personal choice and which one they would avoid given a choice. Figure 4 summarises the results for group members and initiators respectively.

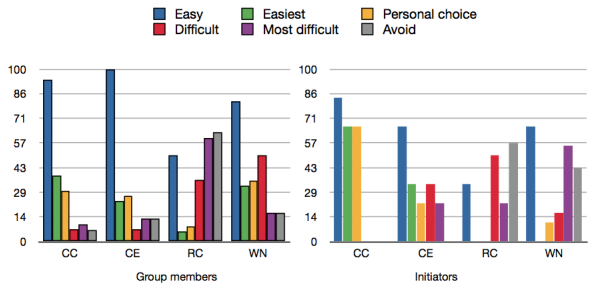

Figure 4: Preferences: Group members (left) and Initiators (right). CC = compare and confirm, CE = copy and enter, RC =repeated comparison, WN = word-matching and number-typing

It is interesting to note that even though the tasks for initiators and non-initiators were different, both graphs follow similar trends; methods that are preferred by group members are also preferred by initiators. It is also surprising to note that even methods that had similar tasks for initiators were rated differently. For example, for CC, CE, and RC initiators only read out a number displayed on their devices. Looking at the data, however, it shows that initiators gave 'holistic' ratings; considered the tasks they performed and also that of group members.

6. DISCUSSION

We have presented a study on the usability and security of four methods for secure device associations in group scenarios. Unlike pairwise associations, completion times in group scenarios are affected by activities of individual group members. These activities tend to normalise completion times and hence minimise the differences between methods.

The main source of failures (both security and usability) is lack of communication between group members and initiators. A failure on a group member's device that is not communicated to initiator may be interpreted as success by initiator. Initiators may also fail to communicate effectively to group members. For example, in one group during the study, initiator read a wrong fingerprint only to realise it after group members had typed it in and the association rejected by devices.

Even though, in this study, initiators correctly observed the number of devices and rejected device association when unexpected number of devices was displayed, it is possible that this may be problematic in day-to-day interactions. A possible solution is having initiators commit to the number

of devices they expect. Initiators may be asked to enter the number of participants before device association which initiator's device can verify once association is complete.

While preferences are subjective and difficult to quantify, they provide an insight into how (potential) users feel about a particular interaction. In our interviews, participants preferred certain methods to others because they felt those methods are either easy to use or secure. This may be a possible explanation to why Figure 4 shows that group members rated CC and CE highest in terms of ease while WN is the most preferred. There are, however, a number of questions that need answers:

1. What is the distance between easy and difficult? To answer this questions requires referring back to our interviews and rating scores. From interviews, it was evident that because questionnaires asked which methods were difficult, participants felt obliged to nominate at least a method. Looking at ASQ scores and overall ratings, methods labelled difficult in preferences have high ASQ scores. Both scores and interviews show that the distance between easy and difficult is not one that can force a user to use one method over another.

2. What do participants mean when they say they can avoid a particular method? The same analysis as the previous question was applied to this question. We found the same result; that participants nominated methods that they can avoid because they felt compelled to do so. A surprise, however, is the consistency—some methods consistently chosen for being difficult or avoidable. For example, most participants nominated *repeated comparison* for both parameters. It was discovered that this was the case because participants had a reference point; *compare and confirm*. Participants felt *compare and confirm* was easy and sufficient hence no need for a similar method that requires them to compare twice.

In addition to results presented above, the study provided insights about group interactions and their impact on security. We highlight some of these below.

6.1 Security through trial and error

One of the five elements of usability is learnability — emphasising design of user manuals that are accessible to users. The basic assumption is that users will take time to read through manuals and understand how a system works. There is sufficient evidence, especially in secure systems, that users will attempt to use a system first and consult a manual only when it is absolutely necessary. For example, during the study a basic background about the system was given and participants were asked whether they had any questions or understood what was required of them. The response was 'We will give it a go and ask when we get stuck'.

A secure system must not depend on correct execution of instruction inside a user manual but must be designed to accommodate 'trial' phase. This is a learning period that users attempt to 'check' how a system works. To accommodate this phase, possibility of security failures must be minimised to an acceptable level if not eliminated.

One approach to mitigate the risks that trial and error may bring to OOB channels is to employ commitment rather than confirmation. An OOB channel should only reveal partial information that a user can use to commit to a final outcome of an association. With incomplete information, a user is only limited to a commitment rather than confirmation. We earlier discussed one example of commitment to group size where initiator know before hand the number of devices (or participants) expected. Using this information, initiator can enter the expected number before device association is initiated so that a final outcome, acceptance or rejection of number of participants, is determined by initiator's device. This way, a user cannot change the outcome after entering the number of devices expected and reduces the chance of a successful attack assuming an attacker is not able to block messages transmitted by one or more member devices.

6.2 Importance of context

During the study, we realised the importance of context of operation in understanding and analysing issues surrounding a system under investigation. Participants had a clear idea what their 'primary' tasks were at each instance. For example, on using a messaging application one participant commented, *"I am a social worker and hold highly confidential discussions about child welfare and I have reservations in using this system in that environment. I prefer face to face and paper based communication which limits where that information can go. This may be just an age issue but that's how I feel about it"*. Context in laboratory studies not only prevent participants from focusing on security tasks, as though they are primary tasks, but also helps in soliciting data that goes beyond the laboratory setting. For example, when asked which method was most user friendly a participant, rather than focusing on the laboratory setting of six participants, commented, *"All these methods are straight forward but I can imagine where there are 50 of you and want to play a game..."* .

Humans are constantly making security decisions [7]; conscious or unconscious. Different people may make different security decisions under similar circumstances. Naturally risk averse people take less risks in the digital world compared to those who are not. Similarly, those who have been exposed to certain risks tend to be risk-averse towards such risks [30]. In device association, changing context means that users are likely to make different decisions in different environments. Ion *et al.* [12] for example, mention that participants in their study indicated that they can change a choice of an OOB channel depending on whether they are interacting with friends or a business colleague, in a private space (such as an office) or in public. A choice of method based on context is also discussed in [4, 15].

6.3 Sum-of-efforts security

It is widely acknowledged that system security is equivalent to the weakest link in the chain [30]. However, Anderson *et al.* [1] have argued that system security may also depend on best effort or sum-of-efforts. Device association for groups is an example of were security depends on a sum of efforts. While initiators were in 'charge' of their groups, group members were observed helping each other. For example, group members recited fingerprints for other members or took the effort to look at another member's device and help that member make the correct decision. Compared to previous studies of single user device association, group effort reduced the number of failures committed.

Design of secure systems where users work as a group, rather than independently, to achieve a common security goal should exploit the principle of sum-of-efforts. A well designed secure system for groups should ensure that the security or insecurity of a system depends on multiple users rather than a single user. In group device association for example, a large group may be split into smaller groups that compare fingerprints and report success or failure as a group rather than as individuals.

6.4 Insecurity of conformity

An interesting observation was made when a participant's device displayed 'Connection failed'. Some participants felt it was their fault that connection failed and hesitated to communicate 'failure' to other members. Initiators, however, consistently asked each group member what message was displayed on a device except in two cases where it resulted in security failures. Wanting to conform to members of a group is a security concern in other secure systems as well. For example, in a study about password Sasse *et al.* [27] found that users were against locking their machines when they moved away from their desks for fear of being seen as paranoid or not trusting fellow workers.

Users' attitudes and behaviour are influenced more powerfully by what they see than what they are told [18]. While conformity is bad for security as discussed above, it is a good characteristic in some scenarios as it is not only applicable to insecure actions but secure ones too. For example, in the study reported here it was observed that some members of a group may not want to carry out a security task or report the result of it but once they realised that other members are doing it they followed suit.

Device association is an interesting and unique case of a security system. Studies on users' compliance to security policies have been conducted for many scenarios but not in a case where a group of participants work together to achieve a common security goal. Further work in this area may highlight the influence of such group settings on individual's method preferences and acceptance.

6.5 Security beyond user interfaces

Users of secure systems will come up with a plausible explanation of how a system protects their assets. This explanation may be based on a number of factors including experience with other systems, observation of the interface, and the context in which they experience the system. During our study, participants came up with various explanation on how the system worked and what should be done to make it more secure. For example, in answering to why he felt *word-matching and number-typing* was more secure than *compare and confirm* a participant responded, '...it is harder to attack three random words from a big dictionary of 9000 words that's not possible to brute force'. While this explanation is plausible, it is not correct in this instance.

Problems arise when users come up with explanations that cause them to behave insecurely. For example, users of peer-to-peer systems end up sharing files they do not wish to share because they do not realise that the system creates shares by default when they think they have to do it themselves [9]. Secure systems should be designed such that their user interfaces closely represent the state of the system and prevent users from engaging into insecure actions no matter how they think the system works. In the study participants thought, based on experience with Bluetooth, passwords, and PIN numbers, that fingerprints should be kept secret among group members. While this is not correct for protocols in the study, it does not result in insecure behaviour.

6.6 Difficult task implies security

Secure systems that are complex and difficult to use are so pervasive that users encounter them in their day to day interactions. As a result, users have come to believe that a secure system that is complex and difficult to use is more secure than a simple and easy to use system. In this study and our previous studies, participants felt *copy and enter* was more secure than *compare and confirm* simply because the former is more difficult than the latter. In addition, some participants were prepared to type numbers longer than 8 digits for applications they considered sensitive.

This belief from users reveals how pervasive complex and difficult to use secure systems are. It seems not enough is being done even after at least 10 years of Human-Computer Interaction Security (HCISec). This view can only be changed by consistently designing secure systems with usability in mind from the onset and evaluating and improving existing systems.

7. CONCLUSION

We have analysed, evaluated and compared methods for transferring fingerprints among devices for the purpose of bootstrapping security in group scenarios. While it has been believed that group settings may be more subject to failures during the association process compared to single user pairwise associations, our findings show the converse to be true. Group members help each other and cover up the weaknesses of struggling members. Comparatively, there is no statistical significance in the differences among methods evaluated in terms of group completions times, ASQ scores (initiators), and overall rating scores (group members). There is, however, statistical significance in individual group member completion times and initiators overall rating scores. Analysis of ASQ and overall scores, preferences and completion times indicate that all studied methods are acceptable in group settings.

Based on participants' feedback and video analysis, we concluded that in group settings security of device association is a function of a sum of efforts rather than weakest link. Data further revealed that users rarely read instructions before using a new system but learn as they 'get on with it'. Users also believe that a secure system must be complex and difficult to use. In addition we realised how contextualising laboratory studies can lead to richer data and responses from participants.

8. ACKNOWLEDGEMENTS

The authors would like to thank Bangdao Chen of Oxford University Computing Laboratory for insightful discussions and help during preparation of this study. We also thank Research in Motion who supported this work by providing mobile devices and partially funding Ronald Kainda. A.W. Roscoe was funded by grants from the US Office of Naval Research.

9. REFERENCES

[1] R. Anderson and T. Moore. The Economics of Information Security. *Science*, 314(5799):610–613, 2006.

[2] D. Balfanz, D. K. Smetters, P. Stewart, and H. C. Wong. Talking to strangers: Authentication in ad-hoc wireless networks. In *In Symposium on Network and Distributed Systems Security (NDSS '02), San Diego, California*, 2002.

[3] M. Čagalj, S. Čapkun, and J. Hubaux. Key agreement in peer-to-peer wireless networks. In *Proceedings of the IEEE (Special Issue on Cryptography and Security)*. IEEE, 2006.

[4] M. K. Chong and H. Gellersen. Classification of spontaneous device association from a usability perspective. In *In Second International Workshop on Security and Privacy in Spontaneous Interaction and Mobile Phone Use (IWSSI/SPMU)*, 2010.

[5] P. DiGioia and P. Dourish. Social navigation as a model for usable security. In *SOUPS '05: Proceedings of the 2005 symposium on Usable privacy and security*, pages 101–108, New York, NY, USA, 2005. ACM.

[6] D. Dolev and A. Yao. On the Security of Public Key Protocols. In *Information Theory, IEEE Transactions on*, volume 29(2), pages 198–208, 1983.

[7] P. Dourish, E. Grinter, J. Delgado de la Flor, and M. Joseph. Security in the wild: user strategies for managing security as an everyday, practical problem. *Personal Ubiquitous Comput.*, 8(6):391–401, 2004.

[8] L. M. Feeney, B. Ahlgren, and A. Westerlund. Demonstration abstract: Spontaneous networking for secure collaborative applications in an infrastructureless environment. In *International conference on pervasive computing (Pervasive 2002)*, 2002.

[9] N. S. Good and A. Krekelberg. Usability and Privacy: A Study of Kazaa P2P File-sharing. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 137–144, New York, NY, USA, 2003. ACM.

[10] M. Goodrich, M. Sirivianos, J. Solis, G. Tsudik, and E. Uzun. Loud and clear: Human-verifiable authentication based on audio. In *Proc. 26th IEEE International Conference on Distributed Computing Systems ICDCS 2006*, pages 10–10, 04–07 July 2006.

[11] B. S. I. Group. Simple Pairing White Paper. www.bluetooth.com/NR/rdonlyres/0A0B3F36-D15F-4470-85A6-F2CCFA26F70F/0/SimplePairing_WP_V10r00.pdf.

[12] I. Ion, M. Langheinrich, and P. Kumaraguru. Influence of User Perception, Security Needs, and Social Factors on Device Pairing Method Choices. In *SOUPS '10: Proceedings of the 5th symposium on Usable privacy and security, to appear*, 2010.

[13] R. Kainda, I. Flechais, and A. Roscoe. Usability and Security of Out-Of-Band Channels in Secure Device Pairing Protocols. In *SOUPS '09: Proceedings of the 5th symposium on Usable privacy and security*, 2009.

[14] R. Kainda, I. Flechais, and A. Roscoe. *Information Security Theory and Practice. Security and Privacy of Pervasive Systems and Smart Devices*, volume 6033 of *WISTP 2010, Lecture Notes in Computer Sciences*, chapter Secure and Usable Out-Of-Band Channels for Ad hoc Mobile Device Interactions, pages 308–315. Spinger, 4 2010.

[15] R. Kainda, I. Flechais, and A. Roscoe. Secure Mobile Ad-hoc Interactions: Reasoning About Out-Of-Band (OOB) Channels. In *In Second International Workshop on Security and Privacy in Spontaneous Interaction and Mobile Phone Use (IWSSI/SPMU)*, 2010.

[16] R. Kainda, I. Flechais, and A. Roscoe. Security and Usability: Analysis and Evaluation. 2010.

[17] A. Kobsa, R. Sonawalla, G. Tsudik, E. Uzun, and Y. Wang. Serial Hook-ups: A Comparative Usability Study of Secure Device Pairing Methods. In *SOUPS '09: Proceedings of the 5th symposium on Usable privacy and security*, 2009.

[18] J. Leach. Improving user security behaviour. *Computers & Security*, 22(8):685 – 692, 2003.

[19] J. R. Lewis. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *Int. J. Hum.-Comput. Interact.*, 7(1):57–78, 1995.

[20] M. Long and D. Durham. Human Perceivable Authentication: An Economical Solution for Security Associations in Short-Distance Wireless Networking. In *ICCCN*, pages 257–264. IEEE, 2007.

[21] R. Mayrhofer and H. Gellersen. Shake well before use: Authentication based on Accelerometer Data. In *Proc. Pervasive 2007: 5th International Conference on Pervasive Computing*, volume 4480 of *LNCS*, pages 144–161. Springer-Verlag, May 2007.

[22] R. Mayrhofer and M. Welch. A human-verifiable authentication protocol using visible laser light. In *ARES '07: Proceedings of the The Second International Conference on Availability, Reliability and Security*, pages 1143–1148, Washington, DC, USA, 2007. IEEE Computer Society.

[23] J. McCune, A. Perrig, and M. Reiter. Seeing-is-Believing: Using Camera Phones for Human-Verifiable Authentication. In *Proc. IEEE Symposium on Security and Privacy*, pages 110–124, 8–11 May 2005.

[24] L. H. Nguyen and A. W. Roscoe. Efficient group authentication protocol based on human interaction. In *Proceedings of the Workshop on Foundation of Computer Security and Automated Reasoning Protocol Security Analysis (FCS-ARSPA)*, pages 9–33, 2006.

[25] A. Perrig and D. Song. Hash visualization: a new technique to improve real-world security. In *International Workshop on Cryptographic Techniques and E-Commerce (CrypTEC '99)*, pages 131–138, 1999.

[26] A. W. Roscoe, S. J. Creese, M. H. Goldsmith, and M. Xiao. Bootstrapping multi-party ad-hoc security. In *Proceedings of SAC 2006*, 2006.

[27] M. A. Sasse, S. Brostoff, and D. Weirich. Transforming the 'weakest link' — a human/computer interaction approach to usable and effective security. *BT Technology Journal*, 19(3):122–131, 2001.

[28] N. Saxena, J.-E. Ekberg, K. Kostiainen, and N. Asokan. Secure Device Pairing based on a Visual Channel (Short Paper). In *SP '06: Proceedings of the*

2006 IEEE Symposium on Security and Privacy, pages 306–313, Washington, DC, USA, 2006. IEEE Computer Society.

[29] N. Saxena, B. Uddin, and V. Jonathan. Universal Device Pairing Using an Auxiliary Device. In *Symposium on Usable Privacy and Security (SOUPS)*, July 2008.

[30] B. Schneier. Biometrics: Truths and fictions. Crypto-Gram Newsletter, August 15, 1998.

[31] B. Schneier. *Secrets & Lies: Digital Security in a Networked World*. John Wiley & Sons, Inc., New York, NY, USA, 2000.

[32] C. Soriente, G. Tsudik, and E. Uzun. BEDA: Button-Enabled Device Association. In *In International Workshop on Security for Spontaneous Interaction (IWSSI)*, 2007.

[33] C. Soriente, G. Tsudik, and E. Uzun. HAPADEP: Human-Assisted Pure Audio Device Pairing. In *ISC '08: Proceedings of the 11th international conference on Information Security*, pages 385–400, Berlin, Heidelberg, 2008. Springer-Verlag.

[34] F. Stajano and R. Anderson. The resurrecting duckling: security issues for ubiquitous computing. *Computer*, 35(4):22–26, April 2002.