An Investigation of the Usability of a Game for Secure Wireless Device Association

Nitesh Saxena
Department of Computer and Information Sciences
University of Alabama at Birmingham
Birmingham, AL 35294
saxena@cis.uab.edu

Alexander Gallego
YieldMo, Inc.
New York, NY 10011
alex@yieldmo.com

Jonathan Voris
Department of Computer Science
Columbia University
New York, NY 10027
jvoris@cs.columbia.edu

ABSTRACT

Securely associating, or "pairing," wireless devices via out-of-band communication channels is a well established approach. Unfortunately, this technique is prone to human errors that lead to security problems such as man-in-the-middle attacks. To address this problem by motivating users, a previous proposal suggested the use of computer games. Games can make the pairing process rewarding, thus potentially improving its usability and security. This paper presents a usability evaluation of a proposed pairing system called "Alice Says" that achieves pairing based on the memory game Simon. The results of our study indicate that users do indeed find the pairing game to be fun, as previously hypothesized. However, the slow execution speed of Alice Says prompts the need for faster game-based approaches to pairing.

1. INTRODUCTION

Connecting devices via wireless communication channels continues to increase in popularity and promises to do so in the future. This popularity is unfortunately accompanied by a rise in security risks. Wireless channels are easy to eavesdrop upon and manipulate. A fundamental security objective is therefore to secure them. The term "pairing" refers to bootstrapping secure communication between two wireless devices in a way that is resistant to eavesdropping and man-in-the middle attacks.

A promising research direction towards solving the pairing problem is to leverage an audio, visual, or tactile out-of-band (OOB) channel that is governed by human users. Unfortunately, pairing turns out to be a daunting problem in practice. Prior work on pairing raises several fundamental usability and security concerns and related research challenges. Most existing pairing methods are based on Short Authenticated String (SAS) protocols [22, 23] that use very short strings that are only 15 bits or so in length. The level of security provided by these methods may therefore not be sufficient for certain applications. Increasing the length of SAS strings, on the other hand, may lead to poor usability because the process will become lengthier. Further, most pairing methods do not offer the theoretical level of security guaranteed by their underlying protocols, as demonstrated in [2]. This is due to the potential these protocols have for human errors to be committed.

1.1 Computer Games for Pairing

In [1], the aforementioned challenges motivated the authors to design a radically different approach to pairing. Their idea was to utilize computer games in order to pair devices. The incentive that this provides to users is fun and entertainment. Since games are a popular form of entertainment, the hypothesis of [1] was that they may improve the security as well as usability of pairing, and therefore help address the challenges outlined above. This is supported by the principle of extrinsically motivated design [21].

While performing security tasks such as pairing, users may not be aware of or care about the impact their actions have on the security or privacy of their devices and data. Due to this lack of engagement in the security process, users may not do their best at the task. To address this issue, [1] proposed the reframing of pairing not as a tedious procedure that puts a costly burden on users, but rather as a playful process that is enjoyable and entertaining to complete. It aimed to transform this security operation from one that users seek to avoid or complete as quickly as possible into one that they relish. As a result, users will be more attentive to and aware of the steps they must follow while executing pairing and will perform better at it. Another important side effect of utilizing a game that was advanced in [1] was that users will be willing to spend more time during the security process, resulting in increased tolerance for such tasks. In the context of pairing, potentially longer OOB strings may thus be used, thus providing a higher level of security.

In essence, by contextualizing a security task as playful rather than a chore, the usability burden and cognitive load it imposes may be reduced. In [1] this was dubbed the *Tom Sawyer Effect* after a well known event in Mark Twain's literary classic, "The Adventures of Tom Sawyer" [15]. In one of the incidents in this novel, the boy Tom is punished by being forced to paint a fence on his day off. To escape his

plight, the clever Tom treats the task as fun rather than resenting it. Upon observing his delight, his friends insist that they be given an opportunity to paint the fence so that they can enjoy it as well. In the same way that Tom convinces his friends to complete what would otherwise be considered an uninteresting job by treating it as a game, the work of [1] sought to persuade users to be attentive during security operations by making these operations enjoyable. Much like Tom's friends, users will aim to achieve precisely the same security goals before and after the addition of playfulness, but might be more inclined to participate due to the perception of fun.

The game-based pairing approach of [1] is an example of a "Game with a Purpose" that addresses lingering problems in usable security [13]. This is because it is not simply a game for its own sake, but rather a form of entertainment that simultaneously achieves a well-defined objective. In this case, the objective is device pairing.

1.2 Our Contributions

In this paper, we set out to investigate whether games can indeed help in extrinsically motivating users during the pairing task and in improving the usability and security of the pairing process. A game intended for pairing two phones, called "Alice Says," was designed in [1]. Alice Says is based on a popular memory game called Simon. It accomplishes the transfer of OOB strings between the two devices. Our contribution is a within-subjects usability study of the Alice Says pairing game compared to a traditional pairing approach based on numeric transfer. Our evaluation is aimed at determining the level of usability and security that is provided by Alice Says.

The results of our study indicate that, overall, Alice Says was considered to be a fun and enjoyable way to pair devices, confirming the hypothesis of [1]. It was also found to be robust to human errors. However, the slow speed of Alice Says relative to pairing based on numeric transfer was found to be a cause for concern. This result prompts the need for faster pairing games or game like approaches. Based on the various insights drawn from our study, we suggest several ways in which the usability of Alice Says can be improved.

The remainder of this paper is organized as follows. First, Section 2 briefly presents the design and implementation of Alice Says as discussed in [1]. This is followed by Section 3, which reports on a usability evaluation of the Alice Says pairing game compared with pairing using numeric transfer. Section 4 reports the results of our study, while Section 5 discusses its results, implications, and the lessons that we learned from it. We review prior device pairing methods in Section 6 before finally drawing conclusions in Section 7.

2. OVERVIEW OF ALICE SAYS

This section briefly reviews the Alice Says game designed for the purpose of device pairing. A more detailed description is provided in [1]. Alice Says is based on the model suggested in [23] in which wireless devices may establish traditional wireless connections and OOB channels. The latter feature modest bandwidths but are physically authenticatable. The devices first execute a SAS protocol [23, 22] over the wireless channel, which results in a short OOB string per appliance. Matching strings imply a successful pairing session, whereas non-matching strings imply that an attack has occurred. OOB strings can be transferred between the devices with a user's assistance, e.g., traditionally by transferring numbers [7]. The devices can compare these values to determine the outcome of pairing. An adversary is allowed to eavesdrop upon the OOB strings in this approach, but he or she cannot modify them.

Figure 1: Hasbro's Simon

When using Alice Says, the transfer of OOB strings is accomplished through a game. The inspiration of this game comes from Hasbro's Simon Says [4], which is depicted in Figure 1. At its core, playing Simon involves nothing more than the short term memorization of audiovisual patterns and thus minimal changes were required to adapt it for use in pairing.

The basic idea of Alice Says is to encode OOB string into a series audiovisual patterns on one device which a user has to copy onto the other device one by one. "00" corresponds to the flashing of the green square, "01" to the red square, "10" to the blue square, and "11" to the yellow square. For example, the string "00011011" will be encoded into a pattern of length four by flashing the green square first, followed by the red square, then the blue square, and finally the yellow square. A user would then select the corresponding squares in the same order on the display of the other device.

Upon initially starting Alice Says, users are provided with a screen that shows them the name of the game with two menu choices: a *single player training mode* and a *two player pairing mode*. A single player mode is provided to allow users an opportunity to unwittingly train themselves to improve their device pairing performance. The two player mode is what actually accomplishes device pairing. It differs from Simon in that the game does not conclude when a mistake is made. It continues until a sufficient number of OOB bits have been relayed between the two devices. This makes the game robust to human errors.

Alice Says preserves the popular aspects of Simon while updating it to a two player mobile device setting. Its user interface is dominated by four large color buttons as was the case with its ancestor. Also intended to mimic the original was the association of a unique tone with each of these keys. A critical aspect of the original game's appeal was the fact that these sounds were designed to be harmonic irrespective of the order in which they were played [24]. To conduct our usability study of Alice Says, two Nokia N97 mobile phones

were used to realize the Alice Says prototype as documented in [1]. A picture of Alice Says setup from our implementation is shown in Figure 2. A clearer image of the game's core user interface, taken from an emulator, is displayed in Figure 3.

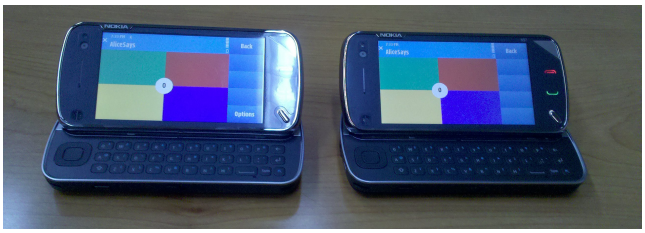

Figure 2: Alice Says Running on Two Nokia N97 Phones

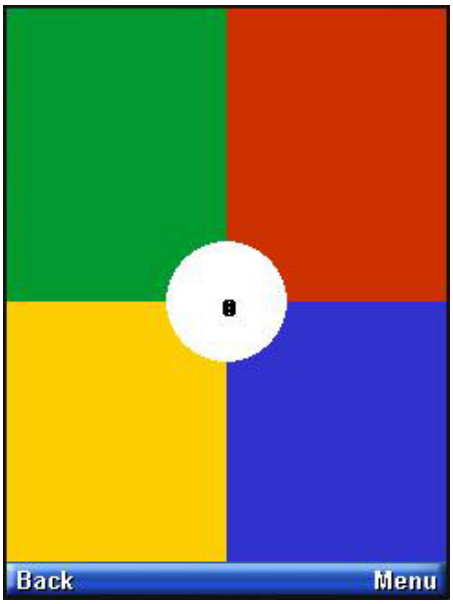

Figure 3: Close Up Image of the Alice Says User Interface

2.1 Example Usage Scenario

The following is an example of an anticipated Alice Says game play pattern [1]. Assume a legitimate pairing session in which a user is consistently able to transfer a pattern of length 5 (i.e., consisting of 10 bits). There are 30 bits in the OOB string that need to be compared.

- In round one, the user will first be provided with a pattern of length one that is just one color or 2 bits. The user will successfully match this pattern. Then the pattern will be extended to length two in round two, and so on.
- Let us say that on the sixth round, the user makes a mistake. At this point the user has successfully transferred the first 10 bits of the OOB string. In the next round the game will begin a new pattern (of length 1 or 2 bits) starting with the 11th and 12th bits of the OOB string.
- Let us say that the user makes another mistake at round 11. Now, the game will begin with a new pattern starting with the 21st and 22nd OOB bits.
- After successfully completing the next 5 rounds, all 30 bits will have been transferred, concluding the game.

3. USABILITY EVALUATION OF ALICE SAYS

3.1 Experimental Framework

A usability study was conducted in order to assess the viability of Alice Says. In order to reduce complexity, we omitted the in-band wireless link between the two mobile devices during our experiments. Since this connection does not have any impact on usability, it was left out in favor of OOB strings created using a pseudorandom pattern. The strings were fixed from subject to subject to prevent some volunteers from receiving strings that were easier to identify than others, but were presented to users in a random order to minimize the effects of learning and fatigue on the test results. Since we employed 30-bit SAS strings and each round of Alice Says encodes two SAS bits, 15 rounds of successful matching were required per pairing session. Our users were also provided with an opportunity to gain some experience using a straightforward numeric transfer mechanism after they tested Alice Says. In order to successfully pair the two devices using numeric pairing, one 30-bit value had to be correctly transferred from one device to the other. This 30-bit number was expressed as 10 decimal digits.

Logging was performed on the N97s to capture the timing of pairing as well as any user mistakes that were made. A feedback interface on a desktop computer was used to present post-conditional questionnaires to users when they completed the study proper. In addition to the System Usability Scale (SUS) [10] queries, the exact phrasing of the survey questions that were presented to volunteers were as follows: 1. The method was enjoyable. 2. The method took a long time. 3. I would like to pair with another user's devices by making use of this method. 4. The sound effects used in this method were pleasant to listen to. 5. I perceive this method to be secure.

Volunteers were also asked to responded to a set of questions regarding the numeric pairing mechanism. These queries were identical to those posed regarding Alice Says with the exception of Question 4. Since there were no sound effects present in the numeric pairing solution, this statement did not apply. Finally, one additional question asked users to explicitly compare the two techniques. The exact wording of this survey item was: I would prefer to use the pairing game rather than use numeric pairing.

3.2 Participant Information

20 test subjects participated in this survey. They were gathered from students, professors, and staff members studying and working in labs at our institution. Word of the study was spread using flyers, emails, and face-to-face recruitment. Ten dollar movie theater gift certificates were offered to testers in order to encourage participation.

Demographic information about these participants was collected during our survey. Half of our subjects were between 18 and 24, while 30% had ages of 25 to 29. There were

several older individuals that were counted in our survey as well. One user was older than 29 but younger than 35 and an additional user was between 35 and 39 years of age. Finally, 10% of our sample were people whose age exceeded 40. A majority, 65%, of subjects were male. One tester did not possess a college degree, 65% had obtained their bachelor's, 25% had obtained their master's degree, and one user had completed his or her doctorate.

The survey also presented users with queries intended to measure their expertise with device pairing and video games. 70% of participants had paired a wireless device before. This was a somewhat surprising result considering the ubiquitous nature of wireless devices. In contrast, every participant responded that they played video games. The fact that more of our user pool was acquainted with video games than device pairing corroborates the assertion that this medium is well suited to address usable security problems.

3.3 Experimental Design

To initiate the experiment, the single player mode of Alice Says was selected. After this, users were presented with the mobile device to provide them with an opportunity to acclimate themselves with the user interface and game play of Alice Says. Once users felt that they had gained enough experience, the formal testing procedure was started by initiating the two player pairing mode. For the first two test cases with each subject, the participant handled the input device while the test administrator took care of the output duties. Thus, to start the game and whenever the volunteer successfully matched a pattern, the administrator selected the "Next" option on the Alice Says menu. In the event that the participant committed a mistake, the test conductor pressed a "previous" button to signal to the device that it should create a new pattern starting with the last incorrectly matched pattern portion.

Meanwhile, the tester's job was to observe the audiovisual pattern displayed on the output device and input it on their appliance. Two test cases were performed in this configuration to ensure that our subjects received generalizable experience. Following this, a third test case was performed with the role of the administrator and subject reversed by switching the mode options on the two phones. This was done in order to give users hands on experience with both sides of the two player game.

When the game-based pairing steps had been completed, users were asked to pair the same set of N97 phones using the numeric pairing technique. For this method, users were asked to transfer SAS data that was expressed as a string of decimal digits between the two devices. To facilitate a fair comparison between the two pairing techniques that were employed, precisely the same test SAS values were used for both operations.

The numeric pairing solution required two modes to function. The first displays the SAS security value, while the second accepted it as input. To begin numeric pairing, one pairing mode was selected on one phone and the opposite setting was selected on the other. Volunteers were then prompted to type the value displayed on the one device on the other that was awaiting their input.

The 30-bit SAS values that we used were twice as long as the 15-bit strings that have typically been used in previous pairing research. To prevent our test participants from becoming overwhelmed by this amount of information, the numeric SAS values were split in half. Initially, only the first portion of the string was shown. The remainder of the SAS data was displayed on a separate screen. In order to transition back and forth between these segments, "next" and "previous" buttons were used on the display device as well.

Users were allowed to move between the two segments as often as they liked and could edit the string they entered on the other phone as often as they wished. No time constraints were enforced during this operation. When finished, test subjects pressed a button on the input device indicating that they were done transferring data. The success or failure of the pairing operation was then indicated to the participant.

Two instances of numeric pairing were executed with each volunteer. In the first case, the test administrator controlled the display device, while the subject manipulated the input on the other. These roles were reversed for the second numeric test run. Testing proceeded in this fashion so we could emulate a social pairing scenario. This was done in order to keep the testing elements as controlled as possible between the two types of device pairing operations.

Following the central portion of the experiment, subjects were presented with a set of post-conditional queries. Beyond the demographic and background information listed in the participant portion, this questionnaire consisted of 30 five point Likert items which were selected to gauge how volunteers felt about Alice Says and numeric pairing. The precise questions posed are provided in the discussion of our testing framework. For both pairing techniques, the first ten of these questions were provided to evaluate the particular technique using the SUS [10].

4. RESULTS

4.1 Efficiency and Errors

For Alice Says, each test subject performed two input sessions and one output session for a total of three test cases per user and 60 overall. The average time that it took to pair using Alice says with a 30-bit transfer was 173.3 seconds with a standard deviation of 28.6 seconds. An average of 1.5 mistakes were made per pairing session. These are partial errors, as pairing completed successfully following their occurrence. Some user errors were caused by an inability to recall the displayed pattern. Others were caused by users accidentally pressing the incorrect color button.

When using numeric pairing, each volunteer completed one input round and one round of output, resulting in two tests per user and 40 altogether. The average pairing execution time for the numeric transfer method was 20.1 seconds with a standard deviation of 9.3 seconds. Remarkably, no errors were committed by any users over the course of the numeric experiments.

4.2 User Feedback

Users awarded Alice Says an average SUS score of 70.5 with a standard deviation of 12.9. For numeric pairing, users

provided an average SUS rating of 78.6 and the standard deviation of these responses was 12.5. We ran an unpaired homoscedastic t-test with a one-tailed distribution on the unaggregated SUS feedback values to see if a substantial distinction existed between our users' opinions of the two pairing techniques. This test resulted in a p-value of 0.028. We can thus conclude with 95% confidence that the difference between the individual SUS ratings of the two solutions that we tested was statistically significant.

Figure 4 presents a graphical depiction of the post-conditional survey responses that we received. When asked if Alice Says was enjoyable, users provided an average response of 3.9. This was the highest aggregate answer that was given to any query posed as a part of this study. The overall answer provided for this question with respect to numeric pairing was a 3.1. Users also responded with a 3.1 regarding the length of the Alice Says pairing procedure. When asked about the length of numeric pairing users provided a mean value of 1.9. In response to the question concerning whether users would like to use the pairing method in question, study participants awarded a 3.8 average to both Alice Says and numeric pairing.

Users agreed that the sounds used in the pairing game were pleasant, awarding this a 3.6 average response. When asked if Alice Says was considered to be secure, users again replied positively with a 3.7 overall. The average response that was provided for the perception of the security of numeric pairing was a marginally higher 3.8. Lastly, when asked if users would prefer to use Alice Says rather than numeric transfer, test subjects responded with a positive average score of 3.7.

Some users also responded to our open-ended question about Alice Says. Several replied in a positive fashion, stating that the paring method was "pretty good," "a fun way to pair," "a very cute game," and that "they had a lot of fun." However, some complained that method was "too long." Some suggested improvements, such as using better menu options than "next and previous," making use of a "better touch screen," and having the game "count with me while I am picking the colors."

The open ended feedback for numeric comparison was varied as well. Some users questioned its security, wondering whether the process could be "compromised using brute force attacks" and characterizing the method as "a little bit simple and not safe." Most volunteers agreed that this method was "very simple and easy to use," though not all, as one individual found it to be "unnecessarily complicated." Though at least one user felt that "no changes are needed," some recommended that the font size of the numeric display be increased. Several users also expressed a lack of engagement in the numeric pairing process by stating that "The numeric pairing method game is not very interesting." The feedback from one user succinctly sums up the numeric feedback we received: "Not as much fun as the game, but definitely faster. I think this might be the method of choice in many situations."

4.3 Effect of Age, Gender and Education

We broke our pool of test subjects down according to several demographic attributes in order to determine if any statistically significant differences existed between these groups. The first characteristic that we analyzed was age. At a 95% confidence level, users under 30 preferred numeric transfer pairing to Alice Says. This mirrors the preference of our volunteers overall, and was expressed as a ten point increase from 71.7 to 81.7 in the average SUS score awarded to these two techniques by users in this group. Though it was not statistically significant, users in the older group did not have much of a preference between the two solutions, as their SUS ratings varied by less than a point from 65.6 for the game-based solution to 66.3 for the numeric alternative. This is perhaps attributable to the fact that younger users have more experience with modern video games and found Alice Says to be less engaging as a result.

Younger users were more forgiving judges of pairing solutions, as they gave both solutions higher SUS scores than the older participants did. The age group comparison for each technique is only statistically relevant at 90% confidence for numeric transfer, however. The p-values associated with our heteroscedastic Student's t-tests for these comparisons are 0.012 for the comparison of Alice Says and numeric transfer for testers under 30, 0.477 for these two methods with testers over 30, 0.212 for the comparison between the two age groups with respect to Alice Says, and 0.068 when the preferences of the two ages groups for numeric pairing were contrasted.

Next, we looked at the subsets of volunteers that were formed by a gender breakdown. Student's t-tests with two-sample unequal variances were performed for these groups as was the done for the comparison between older and younger volunteers. Males awarded Alice Says a 71.5 SUS value and numeric transfer a 78.5. The p-value for this t test was 0.122. Females awarded a higher value, 78.9, to numeric pairing, but a lower value, 68.6, to Alice Says. The comparison between these groupings yielded a p-value of 0.035. This suggests that men liked Alice Says more than women, while women preferred numeric pairing to a greater degree then men did. Both groups favored numeric pairing to the game-based solution, though. The p-values for the t-tests for each pairing solution between the two gender groups were 0.296 for the game-based solution and 0.467 for the numeric pairing process. Based on these p-values, the only gender comparison that was statistically notable was the indication that females preferred numeric pairing, which can be concluded with 95% confidence.

The third and final user group partition that we considered was education. We divided our users into those who had obtained graduate degrees and those who had not. Less educated users gave Alice Says a 71.7 average SUS, while more educated users provided a mean response of 65.6. For numeric transfer, less educated volunteers provided an average SUS score of 81.7 and those with more education gave it a 66.3 on average. Thus less educated users like both methods more, yet education level does not seem to affect people's preference for numeric pairing. Out of these, the comparisons between Alice Says and numeric transfer for both user education groups were found to be significant with a 90% confidence level. This is because their two-sample unequal variance Student's t-tests revealed p-values of 0.094 and 0.077 for less and more educated users, respectively. In

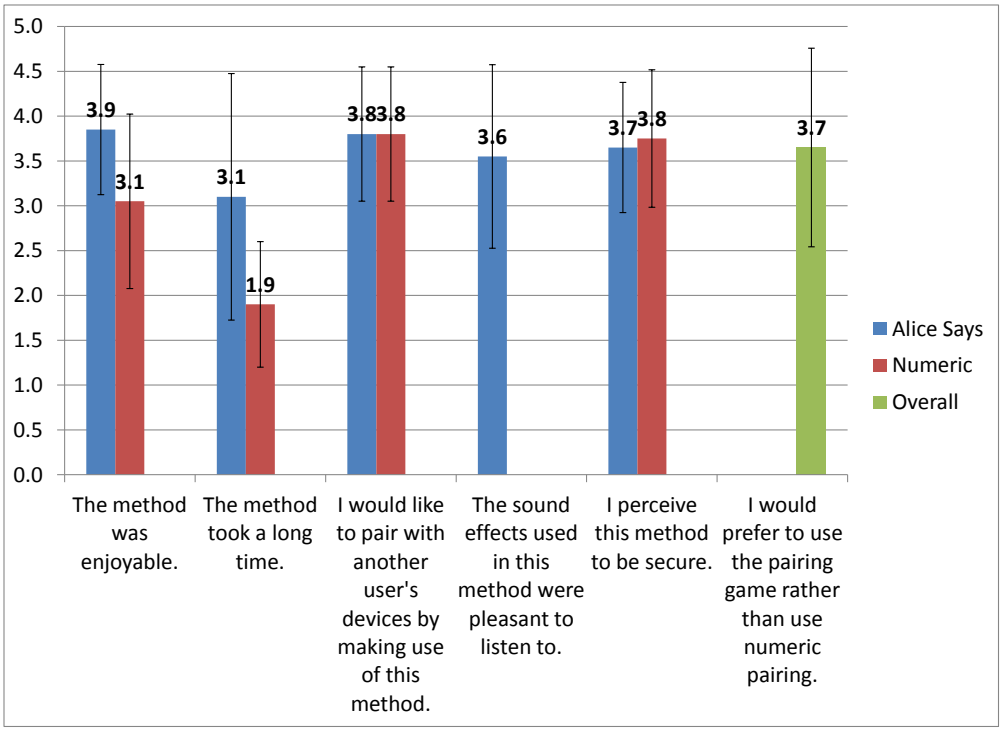

Figure 4: Average Responses to the Post-Conditional Questionnaire

contrast, the t-test performed for Alice Says that compared users by education level returned a p-value of 0.236, while the test regarding numeric pairing for these subsets of users led to a p-value of 0.431.

4.4 Ecological Validity and Study Limitations

This section documents the ways in which this study conformed to and deviated from the real life situation it was intended to capture. A large difference between this study and an actual setting is the absence of a wireless link between the two phones being paired. This may have had an impact on efficiency because the latency this connection would introduce was not taken into account in our study, but this takes negligible time compared to the other steps involved in Alice Says and numeric transfer based pairing.

Another element that should be kept in mind while considering these results is that our test procedure initially asked each user to test Alice Says. Afterwards, they were asked to perform numeric pairing. Their perception of the numeric transfer procedure may therefore have been influenced by their experience with Alice Says. This almost certainly had an influence on their responses to the post-conditional questionnaire. For example, users may not have responded that they felt numeric pairing was an efficient procedure if they did not experience the much longer game-based method shortly beforehand. Similarly, using Alice Says prior to performing numeric pairing may have played a part in the unexpectedly flawless performance which our volunteers demonstrated while using the latter technique.

Beyond these issues, the only other ecological concerns associated with this experiment are those encountered by usability studies in general. For example, by conducting the tests in a lab rather than in their own home, test subjects may have performed with a heightened awareness of their actions. Testers may have altered their responses to survey questions due to a desire to please the investigators, though this effect was guarded against by providing subjects with privacy while answering as well as by anonymizing the results. Finally, providing motivation for the test helped users have a better understanding of how their actions related to real life scenario, but may also have had a conditional effect. A pre-condition questionnaire was not provided in order to minimize this impact, however.

Another usability variable which is difficult to model in a laboratory setting is the level of expertise of the participants involved. Since Alice Says is based on a game with widespread popularity, a majority of individuals will be familiar with its overall steps before using it even for the first time. Furthermore, recall that a single player version of the game was included in our implementation. The inclusion of this variant provides users with an opportunity to play on their own, entertaining themselves and unknowingly honing their pairing acumen along the way. Due to these factors, typical pairing participants are expected to be well acquainted with Alice Says, if not experts in its execution. The prototypical nature of our pairing solution made it impossible to recruit participants with this experience level, however.

We took two steps to compensate for this discrepancy. First, users were allowed to familiarize themselves with the game play of Alice Says prior to using the full two player pairing version. Second, an administrator served as one of the two

users for each test case. While our participant pool may still have been less experienced than the expected real life users of our solution, note that this should have diminished our efficiency and usability results rather than artificially inflating them.

5. DISCUSSION

The data that we collected regarding the suitability of Alice Says as a solution to the problem of device pairing were a decidedly mixed bag. In support of Alice Says, users found game-based pairing to be more enjoyable than numeric pairing and stated that they preferred it. On the other hand, users awarded a higher average SUS score to numeric pairing and felt that the numeric technique was more secure. Further, numeric pairing was faster than the game-based solution by a broad margin.

5.1 Efficiency

The efficiency of Alice Says is clearly its least desirable characteristic. It took slightly under three minutes to complete on average while our numeric pairing solution took just over twenty seconds. Alternative pairing approaches take approximately ten to twenty seconds [2], which is comparable to our numeric technique. Note that we use 30 bits of OOB data, whereas prior studies used 15 to 20 bits. It was hypothesized in [1] that this time frame would be less problematic for a game-based pairing method then it would be for a more traditional technique because users who had a good time while executing the pairing process may wish to extend the technique's execution time rather than reduce it. Contrary to this intuition, users did in fact indicate that the time taken to complete Alice Says was lengthy.

An important lesson learned via our study is the importance of efficiency in device pairing. Since users were enjoying themselves while performing pairing we thought that the speed of the process would allow users to relish the game. This was reflected in the game-based solution for random number generation by Halprin and Naor [19], which did not suffer from any negative usability effects despite taking far longer to complete than traditional techniques. Unfortunately, this turned out not to be the case as users place a high priority on speed in this setting. Our results therefore prompt the design of pairing games that minimize execution time.

5.2 Reliability

While users committed one and half errors per session an average with Alice Says, the pairing game was designed to be resilient in the face of such mistakes. As a result, pairing concluded successfully for all attempts. We expect that the level of partial errors will be reduced as users become more familiar with this approach. Though this was a very promising result, it was outshone by the robustness that was demonstrated by the numeric pairing method that we tested. Absolutely no errors were observed while users were pairing via numeric transfer. This is an unexpected and remarkable result, as all previous user studying involving pairing techniques that are based on numeric values have found these methods to suffer from errors to some degree [7, 17].

Users of Alice Says did not commit any errors that prevented them from successfully achieving pairing. That is, no mistakes were made by participants that would have required them to exit the pairing session and restart from scratch. They also did not experience any errors at all while pairing via numeric transfer. The numeric pairing solution that we tested during the course of this study succeeded in being both wholly error free and efficient in contrast to prior studies [7, 17]. Due to the number of experimental variables involved, we were unable to deduce the precise source of this surprisingly positive result. Possible contributing factors include providing users with ample time, the forward and backwards button, and the precise configuration of the devices that were used. However, we strongly suspect that exposing users to Alice Says prior to performing numeric pairing served as a kind of warm-up exercise that had a substantial influence on their ability to pair devices via numeric transfer. More research is required to substantiate this claim, though.

5.3 User Feedback

Average SUS scores range from 60 and 70 [11]. The Alice Says SUS value of 70.5 should therefore be considered a positive result. The grade of 78.6 that participants awarded numeric transfer is even more desirable. Test subjects agreed with the positive statements regarding Alice Says and numeric pairing. A majority of users concurred that both methods were enjoyable and secure. Unfortunately, users also agreed with the negative statement that was put forth with respect to the timing of Alice Says. In contrast, users did not agree that numeric pairing took too long to accomplish. This discrepancy prompts the need for further work on designing efficient pairing games.

Comparing the feedback that was provided for each question between the two techniques that were tested, users found Alice Says to be more enjoyable, but also found it to be more time consuming. The raw feedback provided by users indicated that they found the game-based pairing process to be entertaining while numeric transfer was perceived as dull. The equivalent values provided regarding whether users would like to use each pairing method indicate a lack of a preference one way or the other. The average score that was awarded to the question regarding the security of numeric transfer was marginally higher for numeric comparison than it was for Alice Says, in contradiction with the open ended feedback we received regarding this pairing technique. Last, but not least, users strongly agreed with the statement that they would favor pairing their devices with Alice Says over the numeric manner of doing so. This conflicts with the SUS scores that users responded with, though.

5.4 Potential Improvements

An interim solution to this issue is to optimize Alice Says in terms of speed by determining an optimal duration for which colors are shown. An item of future work is to experimentally determine the timing threshold that people deem appropriate for a game that is used in a device pairing setting. Broadly speaking, however, the future design of fast pairing games presents a research challenge because most games are not completed in a matter of seconds, but rather minutes or hours. However, there has been a movement towards very brief games played in rapid succession, known as microgames, in the video game community that may be

adaptable as a solution. Nintendo's WarioWare series is a prime example of this [18].

Other possible improvements to Alice Says include adding another counter that "counts with" users as they are inputting the color pattern. Though this would be a stark deviation from the original game design, it was a feature suggested by one of the test subjects as something that would improve usability. Another user suggestion was to use devices with more responsive touch screens. Several complaints were made about how the touch screen on the Nokia N97s had a negative impact of the usability of the game as a whole. A device with a more modern capacitance based touch interface would perhaps not suffer from this drawback.

The efficiency problem is worse in cases where mutual authentication is required; close to 6 minutes of pairing time would be needed with Alice Says, which is not practical. The game can instead be omitted in one direction by having users transfer the result of pairing from one device to the other as suggested in [16]. However, this would make the pairing process vulnerable to errors as well as prone to rushing user behavior [7]. Alice Says can be used to address this issue. For instance, the transmission from device A to B could take place using numeric representations which will be faster than Alice Says. Then the result can be transmitted via a game similar to Alice Says by hiding the result bit within a short random string. This would address both the problem of users being unmotivated and erratic while selecting the correct option on device A as well as that of the potential slowness of the pairing process.

6. RELATED WORK

This section reviews prior traditional pairing methods, that is, those that are not based on games. Stajano and Anderson [8] proposed establishing a shared secret between two devices using a link created through a physical cable. However, in many settings establishing physical contact might not be possible; the devices might not have common interfaces or it might be too cumbersome to carry the cables. Balfanz, et al. [6] extended this approach through the use of an infrared channel. Here devices exchange their public keys over a wireless channel followed by exchanging hashes of their respective public keys over infrared. The main drawback of this technique is that it is only applicable to devices that are equipped with infrared transceivers. Moreover, the infrared channel is not easily perceptible by human users.

Another approach is to perform the key exchange over a wireless channel and authenticate it by requiring that users manually compare the established secret on both devices. Since manually verifying the established secret is cumbersome for users, methods have been designed to simplify it. These include Goldberg's Snowflake mechanism [9] and the Random Arts visual hash [3] by Perrig and Song. These methods require high resolution displays and are thus only applicable to a limited number of devices, such as laptops.

Based on the pairing protocol of Balfanz et al. [6], McCune et al. proposed the "Seeing-is-Believing" (SiB) method [12]. SiB involves establishing two unidirectional visual channels; one device encodes data into a two dimensional barcode and the other device reads it using a camera. Since it requires both devices to have cameras, it is only suitable for pairing devices such as camera phones. Moreover, a recent study [2] shows that users may not be comfortable handling cameras.

Goodrich, et al. [14], proposed "Loud-and-Clear (L&C)", a pairing method based on "MadLib" sentences. This system encodes OOB data into MadLib sentences that users compare on two devices. This method is not applicable to when one of the devices does not have a display or a speaker, however. Saxena et al. proposed a pairing method based on a visual channel [16]. It uses a SAS protocol [22] and is aimed at pairing two devices, A and B, where only B has a relevant receiver.

Uzun et al. [7] carried out a comparative usability study of pairing methods. They consider scenarios where devices have at least 4-digit displays. In what they call the "Compare-and-Confirm" approach, users read and compare SAS data. The "Select-and-Confirm" approach, on the other hand, requires users to select a string on one device that matches with a string on the other device. The third approach, "Copy-and-Confirm," requires that users read data from one device and input it on another.

Recent papers have focused upon pairing devices which lack good interfaces. Access points and headsets are examples of this kind of device. These constraint oriented pairing solutions include the BEDA method [5], which requires that users transfer SAS strings from one device to another using button presses. In [20], Saxena et al. presented a similar pairing method that is universally applicable. It involves users comparing very simple audiovisual patterns such as "beeping" and "blinking."

7. CONCLUSION

This paper presented a usability study of "Alice Says," a system for pairing devices via a game. Alice Says is part of a proposed research direction that involves applying computer games to solve issues in usable security. This experiment was aimed at determining the feasibility of using a game to pair devices in terms of usability and security. Our results indicate that, overall, Alice Says is a fun and an enjoyable way to pair devices, which confirms the previous hypothesis. It was also found to be robust to human mistakes. However, the relatively slow speed of Alice Says pairing compared to other pairing solutions was found to be a cause for concern. This prompts the need for the design of faster pairing games. We intend to develop more efficient alternatives and study the usability and security of these mechanisms as future work.

8. REFERENCES

[1] A. Gallego and N. Saxena and J. Voris. Playful Security: A Computer Game for Secure Wireless Device Pairing. In *The 16th International Computer Games Conference (CGames): AI, Animation, Mobile, Interactive Multimedia, Educational & Serious Games*, 2011.

[2] A. Kumar and N. Saxena and G. Tsudik and E. Uzun. Caveat Emptor: A Comparative Study of Secure Device Pairing Methods. In *Conference on Pervasive Computing and Communications*, 2009.

[3] A. Perrig and D. Song. Hash Visualization: a New Technique to improve Real-World Security. In *Workshop on Cryptographic Techniques and E-Commerce*, 1999.

[4] BoardGameGeek. Simon. Available at `http://www.boardgamegeek.com/boardgame/5749/simon`, 1978.

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

[6] D. Balfanz and D. Smetters and P. Stewart and H. Wong. Talking To Strangers: Authentication in Ad-Hoc Wireless Networks. In *Network and Distributed System Security Symposium*, 2002.

[7] E. Uzun and K. Karvonen and N. Asokan. Usability Analysis of Secure Pairing Methods. In *Usable Security*, 2007.

[8] F. Stajano and R. Anderson. The Resurrecting Duckling: Security Issues for Ad-hoc Wireless Networks. In *Workshop on Security Protocols*, 1999.

[9] I. Goldberg. Visual Key Fingerprint Code. Available at http://www.cs.berkeley.edu/iang/visprint.c, 1996.

[10] J. Brooke. SUS - A quick and dirty usability scale. In *Usability Evaluation in Industry*, 1996.

[11] J. Lewis and J. Sauro. The Factor Structure of the System Usability Scale. In *Conference on Human Centered Design*, 2009.

[12] J. McCune and A. Perrig and M. Reiter. Seeing-Is-Believing: Using Camera Phones for Human-Verifiable Authentication. In *Symposium on Security and Privacy*, 2005.

[13] L. von Ahn. Games with a Purpose. In *Computer Magazine*, 2006.

[14] M. Goodrich and M. Sirivianos and J. Solis and G. Tsudik and E. Uzun. Loud and Clear: Human-Verifiable Authentication Based on Audio. In *Conference on Distributed Computing Systems*, 2006.

[15] M. Twain. The Adventures of Tom Sawyer. 1876.

[16] N. Saxena and J. Ekberg and K. Kostiainen and N. Asokan. Secure Device Pairing based on a Visual Channel. In *Symposium on Security and Privacy*, 2006.

[17] N. Saxena and J. Voris. Pairing Devices with Good Quality Output Interfaces. In *Workshop on Wireless Security and Privacy (WiSP)*, 2008.

[18] Nintendo. WarioWare: Smooth Moves. Available at `http://www.nintendo.com/games/detail/7vgUzwrkjswZ6wiUXTtZQB8ji6_uPB3v`, 2010.

[19] R. Halprin and M. Naor. Games for Extracting Randomness. In *Symposium on Usable Privacy and Security*, 2009.

[20] R. Prasad and N. Saxena. Efficient Device Pairing using "Human-Comparable" Synchronized Audiovisual Patterns. In *Applied Cryptography and Network Security*, 2008.

[21] R. Ryan and E. Deci. Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. In *American Psychologist*, 2000.

[22] S. Laur and K. Nyberg. Efficient Mutual Data Authentication Using Manually Authenticated Strings. In *Conference on Cryptology and Network Security*, 2006.

[23] S. Vaudenay. Secure Communications over Insecure Channels Based on Short Authenticated Strings. In *Crypto*, 2005.

[24] The International Arcade Museum. Touch-Me - Arcade by Atari Games. Available at `http://www.arcade-museum.com/game_detail.php?letter=T&game_id=12694`, 2010.