On the Usability of Secure Association of Wireless Devices Based On Distance Bounding

Mario Cagalj[1], Nitesh Saxena[2] and Ersin Uzun[3]

[1] FESB
University of Split, Croatia
[2] Computer Science and Engineering
Polytechnic Institute of New York University
[3] Information and Computer Sciences
University of California, Irvine
mario.cagalj@fesb.hr, nsaxena@poly.edu, euzun@ics.uci.edu

Abstract. When users wish to establish wireless communication between their devices, the channel needs to be bootstrapped first. Usually, the channel is desired to be authenticated and confidential, in order to mitigate any malicious control of or eavesdropping over the communication. When there is no prior security context, such as, shared secrets, common key servers or public key certificates, device association necessitates some level of user involvement into the process. A wide variety of user-aided security association techniques have been proposed in the past. A promising set of techniques require out-of-band communication between the devices (e.g., auditory, visual, or tactile). The usability evaluation of such techniques has been an active area of research.

In this paper, our focus is on the usability of an alternative method of secure association – *Integrity regions* (I-regions) [40] – based on distance bounding. I-regions achieves secure association by verification of entity proximity through time-to-travel measurements over ultrasonic or radio channels. Security of I-regions crucially relies on the assumption that human users can correctly gauge the distance between two communicating devices. We demonstrate, via a thorough usability study of the I-regions technique and related statistical analysis, that such an assumption does not hold in practice. Our results indicate that I-regions can yield high error rates, undermining its security and usability under common communication scenarios.

Keywords: Authentication, Distance Bounding, Usable Security, Wireless Networks

1 Introduction

Short- and medium-range wireless communication, based on technologies such as Bluetooth and WiFi, is becoming increasingly popular and promises to remain so in the future. With this surge in popularity, come various security risks. Wireless communication channel is easy to eavesdrop upon and to manipulate, and therefore a fundamental security objective is to secure this communication channel. In this paper, we will use the term "pairing" to refer to the operation of bootstrapping secure communication between

two devices connected with a short-range wireless channel. The examples of pairing, from day-to-day life, include pairing of a WiFi laptop and an access point, a Bluetooth keyboard and a desktop.

One of the main challenges in secure device pairing is that, due to sheer diversity of devices and lack of standards, no global security infrastructure exists today and none is likely for the foreseeable future. Consequently, traditional cryptographic means (such as authenticated key exchange protocols) are unsuitable, since unfamiliar devices have no prior security context and no common point of trust.

A number of research directions have been undertaken by the research community to address the problem of pairing of *ad hoc* wireless devices. One valuable and well-established research direction is the use of auxiliary – also referred to as "out-of-band" (OOB) – channels, which are both perceivable and manageable by the human user(s) who own and operate the devices[4]. An OOB channel takes advantage of human sensory capabilities to authenticate human-imperceptible (and hence subject to Man-in-the-Middle or MitM attacks) information exchanged over the wireless channel. OOB channels can be realized using senses such as auditory, visual and tactile. Unlike the in-band (wireless) channel, the attacker can not remain undetected if it actively interferes with the OOB channel. A number of device pairing methods based on a variety of OOB channels have been proposed (we overview these methods later in Section 5; see [18] for a relevant survey). Usability evaluation of these methods is an active research area these days [18, 16, 14].

The focus of this paper is on an alternative approach to device pairing, called *Integrity regions (I-regions)*. I-regions is based on distance bounding [4] and can be implemented using ultrasonic or radio time-of-arrival ranging techniques. It relies on range measurements to prevent MitM attackers from inserting forged messages into the communication between the devices. Basically, the distance bounding technique allows a communicating device A to compute an upper bound of its (physical) distance d from another device it is being paired with. Note that the latter can be device B, with which the user of A intends to pair her device or it could be an MitM attacker. An MitM attack can be effectively foiled if the user controlling A can verify whether the actual distance between A and B is less than or equal to d and make sure that there is no other device (except B) at a distance less than or equal to d. Figure 1 illustrates an MitM attack scenario for I-regions. In the figure, *inter-device distance* denotes the actual physical distance between the two devices (i.e., between the user's phone and kiosk) and *attacker distance bound* is the actual physical distance between user's phone and attacker's device. In this example, attacker distance bound (6 ft) is larger than inter-device distance (3.5 ft), which indicates to the user an ongoing MitM attack. As defined in [40], an *integrity region* is a space centered at user's location, within which the user can confidently establish the presence (or absence) of other wireless devices.

Motivation and Contributions: In this paper, our focus is on the "User Layer" of the I-regions method. In I-regions, once A computes the upper bound of its distance d from B, and shows it on its screen, the user (controlling A) is required to perform two tasks: (1) determine if the perceived distance between A and B is not more than d, and (2)

[4] This has been the subject of recent standardization activities [37].

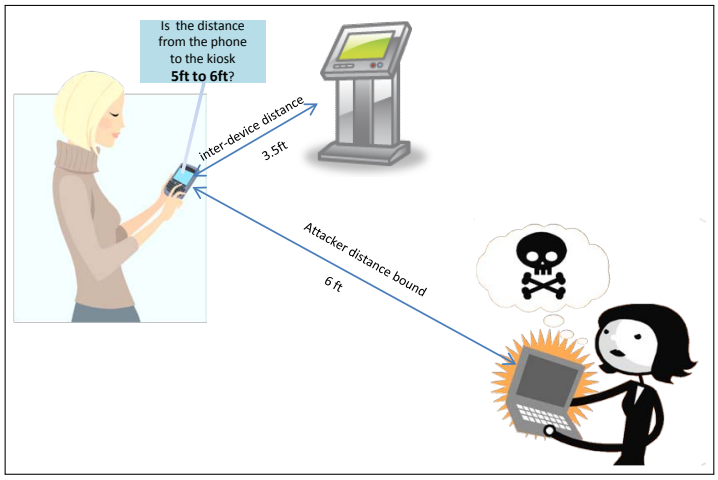


Fig. 1. An MitM Scenario for I-regions (user intends to pair her phone with the kiosk)

make sure if there is no other device (except B) at distance less than or equal to d, i.e., if B belongs to A's integrity region.[5] Clearly, a pre-requisite to the security of I-regions is users' ability of distance judgement. In other words, if users can not correctly perceive the distance shown on devices' screens as well as the distance between the two devices, the security of I-regions can not be guaranteed.

We hypothesize that users perception and interpretation of physical distances (needed to execute the first task mentioned above) is far from accurate. Consequently, I-regions is quite likely to result in both safe errors [38] (i.e, rejection of a valid pairing attempt) and more critically, fatal errors (i.e., acceptance of an MitM attack). In order to test our hypothesis and to evaluate I-regions in terms of efficiency (i.e., speed), robustness (i.e., error tolerance) and usability (i.e., System Usability Score of the method and user's self-confidence about distance judgement), we pursue a thorough and systematic usability study. We remark that such an experimental study was necessary to evaluate I-regions, which is akin to human behavior.

Based on the results of our study, I-regions can be termed quite efficient in terms of completion time. As hypothesized, however, in general (i.e., for arbitrary values of inter-device distance and attacker distance bound), I-regions exhibits poor robustness, with high likelihood of users committing both safe as well as fatal errors. This undermines the security of I-regions, either directly (i.e, in case of fatal errors) or indirectly (i.e, in case of safe errors). Thus, we can conclude that I-regions is not a suitable method for all possible pairing scenarios. However, for some specific values of inter-device

[5] The first manual task can be eliminated if device A is only allowed to accept pairing with devices located within a small, pre-determined distance (e.g., less than 1 meter). This would, however, severely limit the utility of I-regions only to scenarios where devices are in close proximity, and at the same time, damage usability by forcing user to move devices within a certain distance bound, which may not always be possible (such as in case of a wall-mounted access point or when two users are sitting across a long table in a meeting room).

distance (1 ft or 3.5 ft) in conjunction with attacker distance bound (at least 4.5 ft or 7 ft, respectively), I-regions shows reasonable level of robustness and might be acceptable in practice.

Organization: The rest of the paper is organized as follows. In Section 2, we describe the I-regions technique. In Section 3, we discuss our usability study aimed at evaluating I-regions, followed by Section 4, in which we present the results of the study and our analysis. Finally, in Section 5, we overview prior work in the area of wireless device authentication and security association.

2 I-regions

Adversarial Model: The security model for I-regions [40] is as follows. It is assumed that the two entities involved in the communication (A and B) trust each other and are not compromised; otherwise, little can be done. Also, it is assumed that the entities know the (public) protocol parameters. An adversary attacking the I-regions protocol is assumed to have full control on the wireless channel, namely, it can eavesdrop, delay, drop, replay and modify messages. The security notion for I-regions protocol in this setting is adopted from the model of authenticated key agreement due to Canneti and Krawczyk [6]. In this model, a multi-party setting is considered wherein a number of parties simultaneously run multiple/parallel instances of pairing protocols. In practice, however, it is reasonable to assume only two-parties running only a few serial/parallel instances of the pairing protocol. The security model does not consider denial-of-service (DoS) attacks. Note that on wireless channels, explicit attempts to prevent DoS attacks might not be useful because an adversary can simply launch an attack by jamming the wireless signal.

Protocol: The I-regions key exchange protocol, based on Diffie-Hellman (denoted DH-IR), unfolds as shown in Fig. 2. Both Alice and Bob calculate the commitment/opening pairs ((c_A, o_A) and (c_B, o_B)) for messages $m_A \leftarrow 0 \| g^{X_A} \| N_A$ and $m_B \leftarrow 1 \| g^{X_B} \| N_B$, respectively. Here, N_A and N_B are k bit long random strings and "0" and "1" are two public (and fixed) values that are used to break the symmetry and thus prevent a *reflection attack* [21]. In the first two messages, Alice and Bob exchange the commitments c_A and c_B. Then, in the following two messages they open the commitments by sending out o_A and o_B, respectively. It is important to stress that a given party opens his/her commitment only after having received the commitment value from the other party. The first four messages are exchanged over a radio link. Having received the commitment/opening pairs (c_A, o_A) and (c_B, o_B), Alice and Bob open the corresponding commitments and verify that "1" and "0" appear at the beginning of $\widehat{m}_B$ and $\widehat{m}_A$, respectively. If this verification is successful, Alice and Bob generate the authentication strings s_A and s_B. Note that the length of each of these strings is k. The main purpose of the last two messages in the DH-IR protocol is to allow Alice to compare s_A against the authentication string s_B generated by Bob, in a secure way. Thus, Alice sends a k-bit long random string N'_A to Bob and measures the time until she received the response from Bob. Bob responds with $R_B \leftarrow \widehat{N}'_A \oplus s_B$, where the sign hat denotes that the

N'_A as transmitted by Alice may have been altered by the adversary. Alice receives $\widehat{R}_B$, where again the sign hat denotes that R_B as transmitted by Bob may have been altered by the adversary. At the same time, Alice calculates the distance d_A and verifies the corresponding integrity region for the presence of devices other than Bob's device (see Section 2). If this verification is successful, Alice knows that (with a high probability) Bob must have transmitted $\widehat{R}_B$, that is, $\widehat{R}_B = R_B$. Finally, if s_A equals $\widehat{R}_B \oplus N'_A$, Alice notifies Bob and they both accept the messages $\widehat{m}_A$ and $\widehat{m}_B$ (i.e., the corresponding DH public keys) as being authentic. Note that $\widehat{R}_B \oplus N'_A = s_B$ in case no attack takes place.

An adversary against the DH-IR protocol can only succeed with a probability at most 2^{-k}, as long as the commitment scheme used in the protocol is secure. To achieve a high level of security, k can be chosen to be arbitrarily long. For details regarding the security arguments of DH-IR, refer to [40].

Implementation: The DH-IR protocol can be implemented using two techniques: (1) using ultrasonic ranging (US) and (2) using radio (RF) ranging. Both exhibit equal security guarantees, but require different equipment attached to the devices.

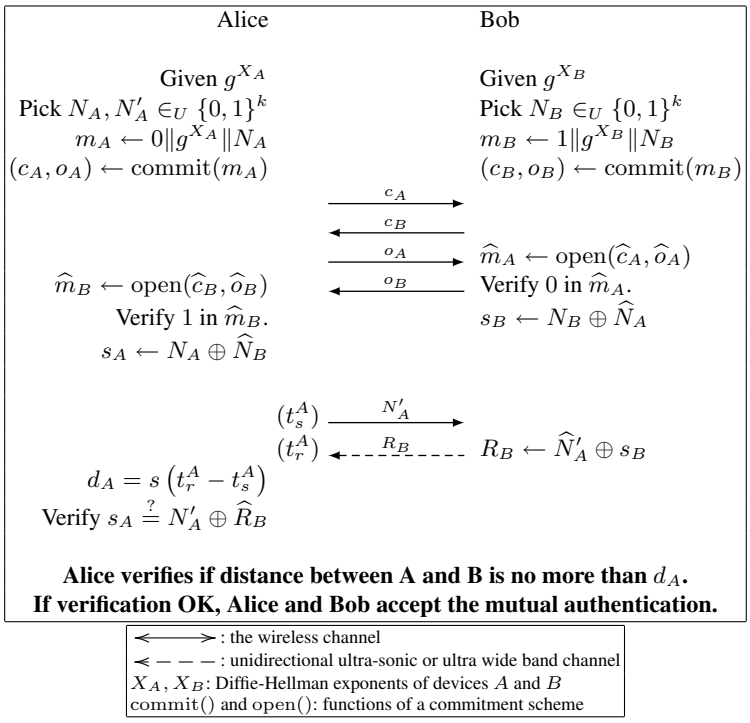


Fig. 2. DH-IR Key Agreement Protocol for I-regions

Ultrasonic ranging requires time measurement precision only in hundreds of μ-seconds, but requires each device to be able to communicate via ultrasonic channel. Current ultrasonic ranging systems (e.g., Cricket motes [29, 27]) can have centimeter precision ranging when the transceivers are perfectly aligned. However, the dependable accuracy is about 0.5ft in practical applications when there are imperfections in the alignment of transceivers.

Radio ranging is more demanding and it requires devices with a high (nanosecond) precision-of-time measurement. To the best of our knowledge, the only commercial technique that achieves such precision, and achieves therefore a high precision-of-distance measurement, is Ultra Wide Band (UWB). In [9], Fontana has demonstrated that with UWB, distances can be measured with an error margin of up to 0.5 ft. Some protocols, e.g., the distance-bounding protocol of Brands and Chaum [4] propose some optimizations through which the cost of nanosecond processing of nodes can be reduced.

In both radio-frequency and ultra-sound solutions, the response time (the XOR operation and the reversion of the transceiver) of the challenged principal must be tightly bound and predictable. With current off-the-self components, ultrasonic ranging seems a more viable implementation of DH-IR and for both techniques the reasonable practical accuracy would be about 0.5ft in a typical use case for I-regions.

3 Usability Evaluation of I-regions

A pre-requisite to the security as well as usability of I-regions is the ability of human users to correctly gauge and interpret the distance between two communicating devices in relation to the distance shown to them as a result of the I-regions protocol.

We hypothesize that human behavior in interpreting distances would be prone to errors. There are two types of possible errors and following the terminology introduced in [38], we call them: (1) safe errors, and (2) fatal errors. Safe errors occur when a user rejects an authentication attempt from an honest device. This happens if the user believes that the distance shown on her device is more than the (perceived) distance between the two devices. As the name suggests, safe errors might not directly undermine the security of I-regions, however, they have an adverse effect on its efficiency and thus usability. Once rejected, the user needs to re-execute another instance of the I-regions protocol by varying the distance between the two devices. This process needs to repeated iteratively until the user has sufficient confidence that she is indeed communicating with the intended device (and not with an attacker). This will clearly slow down the authentication process. In addition, this will lead to poor usability due to user annoyance and increased user burden. Moreover, in certain communication scenarios, it might not be possible to vary the distance between two devices (e.g., two users wanting to communicate in a meeting room). An adversary could also possibly take advantage of such a situation because a user who gets frustrated due to repeated authentication attempts is likely to accept even an attacked session, thus committing a fatal error (which we explain next).

Fatal errors occur when the user accepts an authentication attempt from an attacking device. This can happen if the user believes that the distance shown on her device is less

than or equal to the (perceived) distance between the two intended devices. Fatal errors are clearly dangerous as the user's device will now be communicating with the attacker, even though the user believes her device is communicating with the intended device.

In order to test our hypothesis and to evaluate I-regions, we performed usability experiments. These experiments were simultaneously conducted at two different university campuses: Polytechnic Institute of NYU, USA and University of Split, Croatia.

1. *Efficiency*: time it takes to complete the I-regions method (at the usability layer).
2. *Robustness*: how often the I-regions method leads to safe and fatal errors, with varying inter-device distances.
3. *Usability*: how the method fares in system usability scale [5] and in terms of user confidence in judging distance.

3.1 Testing Apparatus

In our experiments, we used Nokia cell-phones as the testing devices. The models used in the U.S. were N73 and E61 and the model used in Europe was Nokia 6310.[6] We chose to use Nokia cell-phones as they are quite ubiquitous and familiar to many people.

Since our purpose was to test the I-regions method at the usability layer, we chose a simulated test set-up. Our implementation of the I-regions method mock-up was developed to run over the open-source comparative usability testing framework developed by Kostiainen et al. [17] (this framework has previously been used in comparative usability studies of device authentication methods [18]). We used the basic communication primitives as well as automated logging and timing functionalities as provided by this framework.

In terms of user experience, our mock-up closely approximates a real implementation. The two main differences are: (1) our version omits the rounds of the underlying DH-IR protocol, (2) the device only displays the syntactic distance measurement provided by the framework instead of measuring the distance using the packet trip time as in the DH-IR protocol. Notice that the first difference is completely transparent to users as the wireless (and if used, the ultra-sonic) channel is "human-imperceptible." The second difference was necessary to evaluate subjects' ability of comprehending distances and in order to measure resulting error rates.

3.2 Test Cases

We tested the usability of I-regions method with respect to 5 physical distance values, where the actual distances between the devices were set to 1, 2, 3.5, 5 and 6.5 ft. These distances were chosen to capture typical wireless device authentication scenarios. In most situations, the two devices can be within a distance of few feet (e.g., less than 3-4 ft). In some situations, however, it may not be possible to bring the two devices very close to each other (such as in case of a wall-mounted access point or when two users

[6] See `http://europe.nokia.com/phones/n73`, `http://europe.nokia.com/A4142101` and `http://europe.nokia.com/A4143044`, respectively, for the specifications of Nokia phones N73, E61 and 6310.

are sitting across a table in a meeting room). For each inter-device distance value, we created a total of 5 test-cases simulating normal scenarios (i.e., when no attacks occur and the maximum distance shown on device's screen is less than or equal to the physical inter-device distance) as well as attack scenarios (i.e., when a MitM attack is simulated and the maximum distance shown on device's screen is more that the physical inter-device distance). Two of these test cases simulated normal pairing scenarios, while the remaining three simulated attack scenarios, wherein the "attacker distance bound" (i.e., the simulated distance between attacker's device and user's own device) was kept as 1.5, 2.5 and 3.5 ft more than the inter-device distance (Figure 1 depicts an attack scenario). This was done to estimate safe error rates as well as fatal error rates with the attacker residing/hiding within a reasonable proximity of the two devices.

In our study, we only consider the MitM attack cases where the attacker's physical distance is farther than the intended device's. As explained in section 1, users also have to make sure that there is no other device at any distance less than or equal to the actual distance between the intended devices. We did not test users' ability to perform this task. We believe that such attacks, where the attacker is closer than the intended device, would be rare due to higher risk of detection by the user and attacker exposure.

3.3 Test Procedures

In our experiments, all participants were subject to the following procedures (in the given order):

Background Questionnaire: Subjects were asked to fill out a questionnaire through which they were polled for their age, gender and prior experience with device pairing.

Scenario Presentation: Subjects were asked to imagine that they had to send a confidential file from their smart phone to a co-worker's phone. In order to proceed with the file transfer, they needed to first securely pair the two devices.

Experimentation with the Method: Each subject was provided with a test device. The other device was held by the test administrator. The subject was then asked to perform the following procedure a number of times with varying distances between the two devices being paired.

1. Subject was instructed to move to a fixed test point/location previously marked for him/her by the test administrator. He/she was instructed not to move away from this point throughout the experiment.
2. Subject was then given brief and simple instructions on the I-regions pairing method, both textually on the device and orally by the test administrator.
3. After the test administrator set the physical distance between the subject and the administrator's device to one of the pre-defined distances for a given test case, the subject's device showed a (simulated) value for the lower and upper bound distances. Subject then indicated whether the actual physical distance between his device and test administrator's device was within the shown boundaries by pressing the button labeled with his answer.

4. Test administrator relocated the administrator device according to the next test case.

To avoid order effects (particularly due to learning and fatigue), the sequence of test cases was randomized. Also the distance marks/indicators used by the test administrator to correctly set the physical distance according to different test cases were obscured from test participants.

At the beginning of the experiment, we also provided the participants with the choice of distance measurement unit to be used during the experiment. Participants were given the choice of using either the metric (shown in meter and centimeters) or the British units (shown in feet). This was done in order to personalize the pairing method according to individual participants and facilitate better distance comprehension.

In every run of the experiment, the following measures of observable efficiency and robustness indicators were automatically recorded by the testing software: task performance time, fatal errors (if any) and safe errors (if any).

Post-Test Questionnaire: After completing the experiments, subjects completed the System Usability Scale (SUS) questionnaire [5], a widely used and highly reliable 10-item Likert scale that polls subjects' satisfaction with computer systems [2]. We used the original questions from [5], but replaced "system" with "method". Subjects also rated their confidence level on judging physical distances of 1-10 ft with 0.5 ft accuracy, i.e., the typical practical accuracy level provided by distance bounding techniques (as discussed in 2). This allowed us to measure the usability of I-regions as perceived by our participants.

3.4 Subjects

We recruited a total of 43 subjects for our study.[7] 20 of these users participated in our study at the US venue and the remaining 23 of them took part in our study in Croatia. Most participants were students and staff members from the respective universities that we conducted our test at. The subjects were recruited on a first-come-first-serve basis with no controlling or balancing on subject dependent variables such as age and gender. As a result, our sample consisted of a large fraction (85%) of young subjects in the 18-25 age group and a relatively smaller fraction (15%) belonging to the age group of 26-40. We also had a high proportion (70%) of male participants. An overwhelmingly high fraction (91.5%) of our subjects reported prior experience connecting two wireless devices (in response to one of the questions in the Background Questionnaire) and none of them reported any visual disability.

4 Test Results and Interpretations

As we described previously in Section 3.2, each subject participated in a total of 25 test cases (5 test cases each for 5 different values of inter-device distances). Through our tests, we collected data regarding following measures.

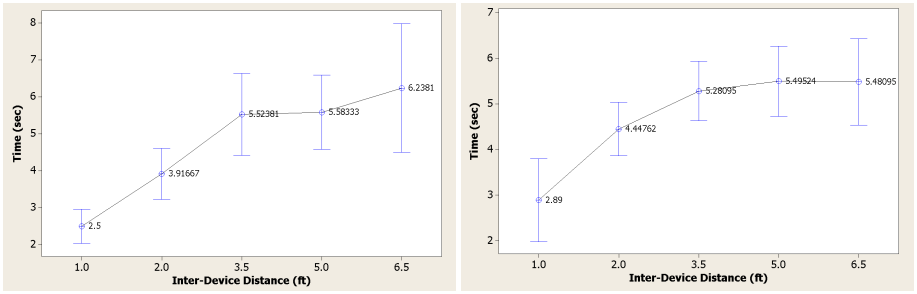
[7] It is well-known that a usability study performed by at least 20 participants captures over 98% of usability related problems [8].

– **Within-subjects measures:** Task performance time, fatal error (categorical) and safe error (categorical).
– **Within-subjects factors:** Test-case varying with respect to (1) physical inter-device distance, (2) the attacker distance bound, and (3) normal or attack scenario.
– **Between-subjects measures:** SUS-score, average task performance time, self-confidence ratings for gauging distances.
– **Between-subjects factors:** Age group, gender and prior experience with wireless device authentication methods.

In the remainder of this section, we discuss our results and interpretations. Unless stated otherwise, statistical significance is reported at the 5% level.

4.1 Overview of Results

Before delving into a detailed analysis of logged data, we provide a brief overview. Following observations were made from the initial analysis of the data collected from 43 subjects, each of whom completed 25 test cases.



(a) Normal (no attack) test cases (b) All test cases

Fig. 3. Task Completion Time vs. Inter-Device Distance

– The mean task completion time, over all test cases, was 4.93 seconds with standard error of 0.84 seconds. Over the test cases simulating normal scenarios, the mean of task completion time was 5.02 seconds with standard error of 0.82 seconds. Figure 3(a) and 3(b) depict the average task completion time for different inter-device distances, calculated over normal test scenarios and over all test scenarios, respectively. When compared to the completion times for other pairing methods studied in [18, 16], we find that I-regions is quite fast for all inter-device distances.
– Depending on the inter-device distance and simulated attacker distance bound, observed fatal error rate ranged from 9.5% to 78.5%. Over all executed test-cases, the observed fatal error rate was 42%. Figure 4 shows the average rate of fatal errors for different test cases. These numbers are alarmingly high, especially when the simulated bound for the attacker distance is close to the inter-device distance, i.e.,

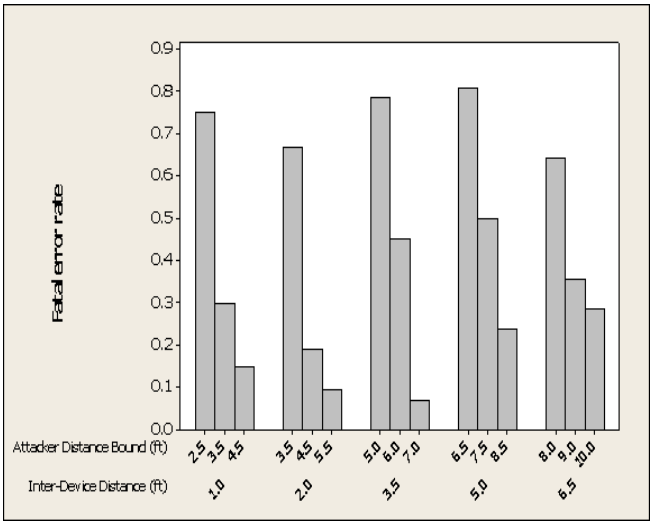


Fig. 4. Fatal Error Rates for Different Test Cases

1.5 ft and 2.5 ft more than the inter-device distance. Recall that a fatal error leads to a successful MitM attack.

- Over all normal test scenarios, the observed rate of safe errors was 29%. Figure 5 shows the observed average safe error rates for different inter-device distances. Although safe error rates are smaller than fatal errors rates (as observed above), they are still quite high (more than 10% in all cases). We believe that safe error rates higher than 10% are problematic since such errors undermine the usability, can cause user frustration and eventually lead to fatal errors.
- The mean SUS-score assigned by the subjects was 75 (out of 100) with standard deviation of 12.4. In general, this means that our subjects were reasonably happy with the method and felt that it is easy to use. This can be seen as a positive indication.
- The mean of all participant responses to the last question of the post-test questionnaire, i.e., the self-confidence level in guessing short distances with 0.5 ft accuracy, was 3.24 with standard deviation 1.14. This was rated on a 5-point Likert scale (1=strongly disagree, through 5=strongly agree). This implies that most of our users believed that their comprehension of physical distances was quite up to the mark.

4.2 Within-Subjects Analysis

We analyze the effect of test case on the efficiency and robustness of I-regions. Repeated measures analysis of variance and Chi-square tests revealed that the type of test case has a highly significant effect on task completion time as well as fatal and safe error rates. To better understand the effect of each independent variable of every test case, we look at them individually.

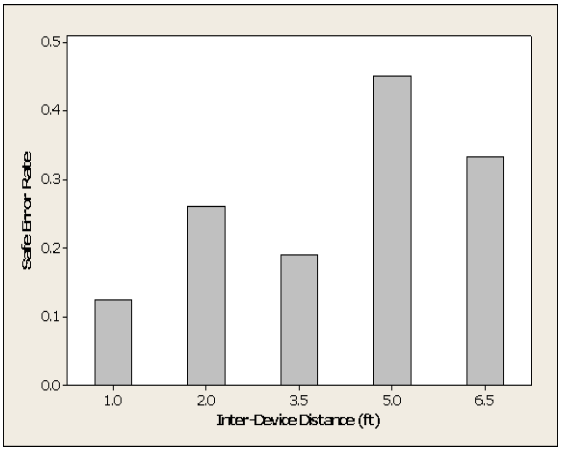


Fig. 5. Safe Error Rate vs. Inter-Device Distance

Physical inter-device distance: Analysis of variance revealed that actual physical distance between devices has a significant effect on task completion time. As shown in Figure 3(a), the mean task completion time, for normal scenarios, gradually increases from 2.50 seconds to 6.24 seconds as the inter-device distance increases from 1 ft to 6.5 ft. A similar pattern can be observed in Figure 3(b) for completion time over all test cases. This finding is intuitive as it is easier (and thus faster) to gauge the distance between devices that are closer compared to those that are farther.

Chi-square tests revealed that inter-device distance also has an effect on the likelihood of committing fatal and safe errors. However, the error rates are not directly correlated with inter-device distance as it was in the case of task completion time.

Normal vs. attack scenario: We did not find any significant difference between the mean completion timings of test cases corresponding to normal and attack scenarios.

Simulated attacker distance bound: The difference between the attacker's distance bound and the inter-device distance has a highly significant effect on fatal errors. As the difference increases from 1.5 ft to 3.5ft, the mean proportion of fatal errors drops from 0.73 to 0.17. This pattern can also be observed in Figure 4, irrespective of the inter-device distance. As expected, this means that fatal errors are more likely to occur when the attacker's distance is close to the actual inter-device distance. In other words, human subjects are expected to be less erratic in detecting a distant attacker.

Simulated attacker distance, on the other hand, did not have any significant effect on task completion time.

4.3 Between-Subject Analysis

Effect of gender: In conducted unpaired t-tests, we have not found any significant effect of gender on task completion time, and users' SUS-scores and self-confidence ratings.

According to the results of Chi-square tests, the effect of gender on fatal error rate and safe error rate were also not significant.

Effect of age: Our test sample consisted of subjects belonging to two age groups, namely 18-25 years and 26-40 years. Unpaired t-tests revealed that subjects belonging to the age group 26-40 take significantly longer time to complete the task compared to the subjects belonging to the 18-25 group (p=0.037). The means of task completion times were 7.128 and 4.744 seconds, respectively, corresponding to the two age groups. There was no significant effect of age, however, on users' SUS-scores and self-confidence ratings.

Chi-Square tests revealed that age has significant effect on fatal (p=0.041) and safe errors (p=0.038). The rate of making a fatal error was 0.44 for the 18-25 age group and 0.32 for the 26-40 age group. Similarly, the rate of safe errors in 18-25 and 26-40 age groups were 0.31 and 0.18 respectively.

A plausible explanation of the above findings is that our slightly older subjects were more conscious while completing the assigned tasks compared to their younger counterparts. This also helps to explain the higher task completion durations for the older group. Another possible reason could be that older subjects were perhaps more familiar with distance measurements.

Self-confidence in distance judgement: ANOVA tests revealed that participants' self-confidence ratings for accurately judging distance have a significant effect on task completion time. Although there was no obvious linear correlation between self-confidence and the completion, we observed that people with the highest confidence ratings tend to have shorter completion times. People with mid-range confidence ratings took the longest and subjects having the lowest confidence had the highest variance in completion times.

Self-confidence ratings also had a significant effect on fatal and safe errors. In both cases, the proportion of errors (Y-axis) are almost *bell shaped* with respect to the self-confidence ratings (X-axis). However, the variance is observed to be higher for subjects with low confidence ratings. For fatal errors, the mean proportions corresponding to self-confidence levels 1, 3 and 5 were 0.29, 0.49 and 0.34 respectively. For the same self-confidence levels, the respective corresponding mean safe-error proportions were 0.18, 0.30 and 0.11.

When we look at the error rates, the most surprising finding was that the mean error rates for subjects with lowest confidence ratings was smaller than the mean error rates for subjects with mid-range confidence ratings. Although hard to explain, this finding could be partly because some subjects rated themselves higher due to optimism and overconfidence biases. On the other hand, some subjects might also have become over-cautious while answering this question and under-rated their confidence level or performed better than they would normally do due to the observer effect (also known as the Hawthorne Effect [19]. This also helps to explain the higher variance observed within the task completion timings corresponding to the group of subjects with lowest self-confidence ratings.

4.4 Discussion of Combined Measures

A usable wireless device authentication method should perform well in terms of all three (not just one of the) measures, i.e., efficiency (task completion time), robustness (likelihood of committing safe and fatal errors) and usability (user ratings and self-confidence). As our analysis in prior subsections indicate, I-regions is certainly quite efficient and can be considered usable in terms of its SUS-score. However, *in general*, I-regions has poor robustness, with high likelihood of safe as well as fatal errors. This means that I-regions might not be a practical method for arbitrary values of inter-device distance and attacker distance bound.

On the other hand, since I-regions exhibited, in spite of its manual nature, quite low task completion time and good usability ratings from the participants, we set out to further explore it. We were interested in investigating whether I-regions is robust (for practical purposes) for any specific values for inter-device distance and attacker distance bound. Looking at Figure 4, we find that fatal error rates are on a lower side (less than 10%), when the distance between attacker's device and user's device is 3.5 ft more than the inter-device distance, especially for inter-device distances of 2.0 ft and 3.5 ft. As mentioned previously, an error rate around 10% might be acceptable in practice for certain scenarios. Similarly, looking at Figure 5, mean safe error rate for inter-device distance of 1.0 ft is 12.5%, which might also be an acceptable fraction in practice, especially for scenarios where user can vary (reduce) the inter-device distance prior to re-executing the authentication process in case a safe error occurs.

We wanted to determine values (if any) of inter-device distance and attacker distance bound optimal with respect to safe errors, fatal errors and task completion time (all taken together). To this end, we first set out to check whether our efficiency, robustness and usability measures were independent of one another. Table 1 shows the correlation coefficients and their respective statistical significance (P-values). As shown in the table, none of the measures is sufficiently correlated with others that it could be justifiably omitted. However, it should be noted that the fatal and the safe errors are positively correlated. Although this correlation is modest, it still suggests that the subjects who accepted incorrect distances showed a tendency to reject correct distances (or vice versa).

	Average Task Performance Time	**SUS-Score**	**Fatal Error Rate**
SUS-Score	-0.242 (0.129)	-	-
Fatal Error Rate	-0.182 (0.249)	0.111 (0.484)	-
Safe Error Rate	0.144 (0.365)	-0.157 (0.332)	0.393 (0.010)

Pearson correlation coefficient
(P-Value)

Table 1. Cross-Correlation of Different Measures

In terms of efficiency, shorter inter-device distances results in better completion times. However, I-regions was quite efficient in general and completion times were almost always under 8 seconds. Compared to other pairing methods, the time required for I-regions is quite low and completion time differences among various inter-device distances were small. Thus, it is more appropriate to concentrate on the combined effect of fatal and safe errors on I-regions. Figure 6 shows this effect for varying inter-device distances and attacker distance bounds. Clearly, the distance values lying on the lower left are considered better. We can observe that although [inter-device distance, attacker distance bound] values [1 ft, 2.5 ft] and [1 ft, 3.5 ft] have low safe error rates, they yield quite high fatal error rates and are thus not suitable. [1 ft, 4.5 ft] and [3.5 ft, 7ft] are the only tuples with reasonable safe and fatal error rates (although such rates may be still high for many practical applications). We can conclude, therefore, that inter-device distance of 1 ft and 3.5 ft, with the attacker distance bound no less than 4.5 ft and 7 ft, respectively, works the best for I-regions. These values might be suitable for certain pairing scenarios. However, for all other values, I-regions can be deemed impractical.

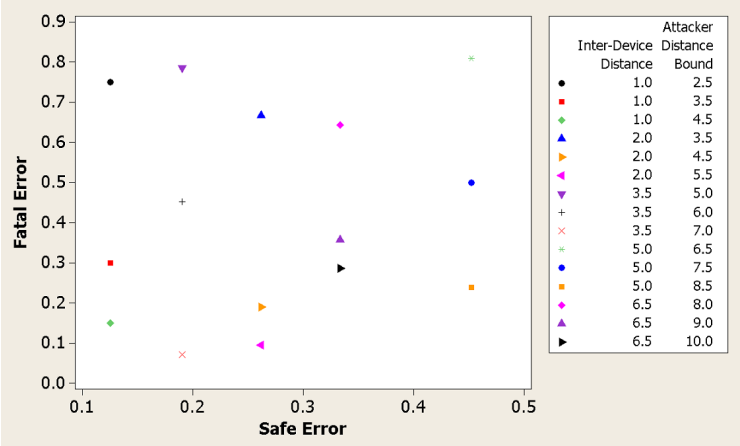


Fig. 6. Mean Fatal Error Rate vs. Mean Safe Error Rate for Different Test Cases

4.5 Summary of Results

Our (significant) findings can be summarized as follows:

- I-regions exhibits low task completion timings and rated as usable by the test participants.
- Most users were quite confident about their distance judgement ability.
- In general (i.e., for arbitrary values of inter-device distance and attacker distance bound), I-regions shows poor robustness, with high likelihood of users committing

both safe as well as fatal errors. However, for some specific values of inter-device distance (1 ft and 3.5 ft) in conjunction with attacker distance bound (at least 4.5 ft and 7 ft), I-regions shows reasonable level of robustness and might be acceptable in practice.
- Fatal errors become less likely as the difference between attacker's distance bound and the inter-device distance increases.
- The task completion time has a tendency to increase as the inter-device distance increases.
- Older subjects (26-40 age group) commit less fatal and safe errors compared to their younger counterparts (18-25 age group). Older subject, on the other hand, took longer to complete the tasks.
- Subjects who felt most confident about their distance judgment abilities (i.e., those with a rating 5) committed less safe errors and completed the tasks faster compared to those having mid-range confidence levels (i.e., those with ratings 2, 3 and 4).

5 Related Work

Providing integrity and authentication over insecure wireless channels is an active area of research. This provision has mainly focused on the key establishment after which the integrity and the authenticity of the messages is ensured by the use of known cryptographic techniques. We review prior work in this area, in the chronological order of publication.

In this context, Stajano and Anderson propose the *resurrecting duckling* security policy model, [36] and [35], in which key establishment is based on the physical contact between communicating parties (their PDAs). In [1], the authors go one step further and relax the requirement that the location limited channel has to be secure against passive eavesdropping; they introduce the notion of a *location-limited channel* (e.g., an infrared link), which is used to exchange pre-authentication data and should be resistant to active attacks.

Another early approach involves image comparison. It encodes a small checksum data calculated over the exchanged data into images and asks the user to compare them on two devices. Prominent examples include "Snowflake" [10], "Random Arts Visual Hash" [26] and "Colorful Flag" [7]. Such methods, however, require both devices to have displays with sufficiently high resolution. A more practical approach, based on SAS protocols [25, 20], suitable for simpler displays and LEDs has been investigated in [30] and [28].

More recent work [24] proposed the "Seeing-is-Believing" (SiB) pairing method. In SiB one device encodes a checksum data into a two-dimensional barcode which it displays on its screen and the other device "reads it" using a photo camera, operated by the user. For bidirectional authentication, the same procedure is executed once more with devices changing roles. A related approach has been explored in [31]. Like SiB, it uses the visual out-of-Band (OOB) channel but requires one device to have a continuous visual receiver, e.g., a light detector or a video camera. The other device must have at least one LED. The LED-equipped device transmits OOB data via blinking while the other receives it by recording the transmission and extracting information based on

inter-blink gaps. The receiver device indicates success/failure to the user who, in turn, informs the other to accept or abort.

Another recent method is "Loud-and-Clear" (L&C) [11]. It uses the audio (acoustic) OOB channel along with vocalized MadLib sentences which represent the digest of information exchanged over the main wireless channel. There are two L&C variants: "Display-Speaker" and "Speaker-Speaker". In the latter the user compares two vocalized sentences and in the former – displayed sentence with its vocalized counterpart. Some follow-on work (HAPADEP [34, 12]) considered pairing devices using only the audio channel. HAPADEP transmits cryptographic protocol messages over audio and requires the user to merely monitor device interaction for any extraneous interference.

Yet another approach: "Button-Enabled Device Authentication (BEDA)" [33, 32] suggests pairing devices with the help of user button presses, thus utilizing the tactile OOB channel. This method has several variants: "LED-Button", "Beep-Button", "Vibration-Button" and "Button-Button". In the first two variants, based on the SAS protocol variant [31], the sending device blinks its LED (or vibrates or beeps) and the user presses a button on the receiving device. Each 3-bit block of the SAS string is encoded as the delay between consecutive blinks (or vibrations). As the sending device blinks (or vibrates), the user presses the button on the other device thereby transmitting the SAS from one device to another. In the Button-Button variant, the user simultaneously presses buttons on both devices and random user-controlled inter-button-press delays are used as a means of establishing a common secret using a password based key agreement protocol (e.g., [3]).

There are also other methods which require hardware that is less common. To briefly summarize a few. [15] suggested using ultrasound and [23] suggested using laser as the OOB channel. A very different OOB channel was considered in "Smart-Its-Friends" [13]: a common movement pattern is used to communicate a shared secret to both devices as they are shaken together by the user. A similar approach is taken in "Shake Well Before Use" [22].

A closely related approach to the method tested in this paper is introduced in [39]. Although practical for establishing secure connection between devices that are in very close proximity, [39] lacks the flexibility to accommodate various distances between devices. This limitation is due to the fact that it uses the environmental radio signal noise as the initiating shared secret between devices and the sensed noise is sufficiently similar only within close proximity. Moreover, the security of using radio noise as a location dependent secret is not well studied and currently unknown at best.

An experimental investigation [38] presented the results of a comparative usability study of simple pairing methods for devices with displays capable of showing a few digits. In the "Compare-and-Confirm" approach, the user simply compares two 4-, 6- or 8-digit numbers displayed by devices. In the "Select-and-Confirm" approach, one device displays to the user a set of (4-, 6- or 8-digit) numbers, the user selects the one that matches the number displayed by the other device. In the "Copy-and-Confirm" approach, the user copies a number from one device to the other. The last variant is "Choose-and-Enter" which asks the user to pick a "random" 4-to-8-digit number and enter it into both devices. All methods except "Choose-and-Enter" are based on SAS protocols and the latter is based on password based key agreement protocols e.g., [3].

Quite recently, more comprehensive studies of different pairing methods have been introduced in [18, 16] and [14]. In [18], authors selected 13 pairing methods that they deem practical and comparatively investigated the security and usability of them. [16, 14] also conducted similar studies but their main focus was usability rather than security. Unfortunately, distance bounding based pairing methods were not included into any of these studies and the usability of such methods left unknown. In this paper, we try to fill this gap left by the previous work and shed light on the usability of distance bounding based pairing methods.

6 Conclusion

In this paper, we presented the results of the first usability study of the I-regions technique. Based on our results, I-regions can be termed quite efficient and it is found to be usable by our subjects. However, in general (i.e., for arbitrary values of inter-device distance and attacker distance bound), I-regions exhibits poor robustness, with high likelihood of users committing both safe as well as fatal errors. This undermines the security of I-regions, either directly (i.e, in case of fatal errors) or indirectly (i.e, in case of safe errors). Thus, we can conclude that I-regions is not a suitable method for all communication scenarios. However, for some specific values of inter-device distance (1 ft or 3.5 ft) in conjunction with attacker distance bound (at least 4.5 ft or 7 ft), I-regions shows reasonable level of robustness and might be acceptable in practice.

Acknowledgements

We are thankful to Arun Kumar and Toni Perkovic for administering our usability studies at our US and Croatian locations, respectively. We also thank Yang Wang for his comments on an earlier version of this paper and CANS'09 anonymous reviewers for their feedback.

References

1. D. Balfanz, D. Smetters, P. Stewart, and H. Wong. Talking to Strangers: Authentication in Ad-Hoc Wireless Networks. In *Proceedings of the 9th Annual Network and Distributed System Security Symposium (NDSS)*, 2002.
2. Aaron Bangor, Philip T. Kortum, and James T. Miller. An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, 24(6):574–594, 2008.
3. V. Boyko, P. MacKenzie, and S. Patel. Provably Secure Password-Authenticated Key Exchange Using Diffie-Hellman. In *Advances in Cryptology-Eurocrypt*, pages 156–171. Springer, 2000.
4. S. Brands and D. Chaum. Distance-bounding protocols. In *Workshop on the theory and application of cryptographic techniques on Advances in cryptology*, pages 344–359. Springer-Verlag New York, Inc., 1994.
5. John Brooke. SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, and A. L. McClelland, editors, *Usability Evaluation in Industry*. Taylor and Francis, London, 1996.

6. Ran Canetti and Hugo Krawczyk. Analysis of key-exchange protocols and their use for building secure channels. In *Advances in Cryptography-Eurocrypt*, pages 453–474, 2001.
7. Carl M. Ellison and Steve Dohrmann. Public-key support for group collaboration. *ACM Transactions on Information and System Security*, 6(4):547–565, 2003.
8. L. Faulkner. Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments, & Computers*, 35(3):379–383, 2003.
9. R.J. Fontana. Experimental Results from an Ultra Wideband Precision Geolocation System. *Ultra-Wideband, Short-Pulse Electromagnetics*, May 2000.
10. Ian Goldberg. Visual Key Fingerprint Code. `http://www.cs.berkeley.edu/iang/visprint.c`, 1996.
11. M. Goodrich et al. Loud and Clear: Human-Verifiable Authentication Based on Audio. In *International Conference on Distributed Computing Systems*, 2006.
12. M.T. Goodrich et al. Using audio in secure device pairing. *International Journal of Security and Networks*, 4(1):57–68, 2009.
13. Lars Erik Holmquist et al. Smart-its friends: A technique for users to easily establish connections between smart artefacts. In *Ubiquitous Computing (UbiComp)*, pages 116–122, London, UK, 2001. Springer-Verlag.
14. Ronald Kainda et al. Usability and security of out-of-band channels in secure device pairing protocols. In *Symposium On Usable Privacy and Security (SOUPS)*, 2009.
15. T. Kindberg and K. Zhang. Validating and securing spontaneous associations between wireless devices. In *Information Security Conference*, pages 44–53, 2003.
16. Alfred Kobsa et al. Serial hook-ups: A comparative usability study of secure device pairing methods. In *Symposium On Usable Privacy and Security (SOUPS)*, 2009.
17. Kari Kostiainen and Ersin Uzun. Framework for comparative usability testing of distributed applications. In *Security User Studies: Methodologies and Best Practices Workshop*, 2007.
18. Arun Kumar et al. Caveat Emptor: A Comparative Study of Secure Device Pairing Methods. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2009.
19. H.A. Landsberger. *Hawthorne revisited*. Cornell University Press, 1968.
20. S. Laur and K. Nyberg. Efficient mutual data authentication using manually authenticated strings. In *International Conference on Cryptology and Network Security (CANS)*, volume 4301, pages 90–107. Springer, 2006.
21. Wenbo Mao. *Modern Cryptography, Theory & Practice*. Prentice Hall PTR, 2004.
22. R. Mayrhofer and H. Gellersen. Shake Well Before Use: Authentication Based on Accelerometer Data. In *Pervasive Computing (PERVASIVE)*. Springer, 2007.
23. R. Mayrhofer and M. Welch. A Human-Verifiable Authentication Protocol Using Visible Laser Light. In *International Conference on Availability, Reliability and Security (ARES)*, pages 1143–1148, 2007.
24. Jonathan M. McCune, Adrian Perrig, and Michael K. Reiter. Seeing-is-believing: Using camera phones for human-verifiable authentication. In *IEEE Symposium on Security and Privacy*, 2005.
25. Sylvain Pasini and Serge Vaudenay. SAS-Based Authenticated Key Agreement. In *Public key cryptography (PKC)*, 2006.
26. Adrian Perrig and Dawn Song. Hash visualization: a new technique to improve real-world security. In *International Workshop on Cryptographic Techniques and E-Commerce*, 1999.
27. H. Piontek, M. Seyffer, and J. Kaiser. Improving the accuracy of ultrasound-based localisation systems. *Personal and Ubiquitous Computing*, 11(6):439–449, 2007.
28. Ramnath Prasad and Nitesh Saxena. Efficient device pairing using "human-comparable" synchronized audiovisual patterns. In *Conference on Applied Cryptography and Network Security (ACNS)*, June 2008.

29. N. B. Priyantha, A. Chakraborty, and H. Balakrishnan. The Cricket location-support system. In *Proceedings of the ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, pages 32–43. ACM Press, 2000.
30. V. Roth et al. Simple and effective defense against evil twin access points. In *ACM conference on Wireless network security (WISEC)*, pages 220–235, 2008.
31. Nitesh Saxena et al. Extended abstract: Secure device pairing based on a visual channel. In *IEEE Symposium on Security and Privacy*, 2006.
32. C. Soriente, G. Tsudik, and E. Uzun. Secure pairing of interface constrained devices. *International Journal of Security and Networks*, 4(1):17–26, 2009.
33. Claudio Soriente, Gene Tsudik, and Ersin Uzun. BEDA: Button-Enabled Device Association. In *International Workshop on Security for Spontaneous Interaction (IWSSI), UbiComp Workshop Proceedings*, 2007.
34. Claudio Soriente, Gene Tsudik, and Ersin Uzun. HAPADEP: human-assisted pure audio device pairing. In *Information Security*, pages 385–400, 2008.
35. F. Stajano and R. Anderson. The Resurrecting Duckling: Security Issues for Ad-hoc Wireless Networks. In *International Workshop on Security Protocols*, 1999.
36. Frank Stajano. *Security for Ubiquitous Computing*. John Wiley & Sons, Ltd., 2002.
37. Jani Suomalainen, Jukka Valkonen, and N. Asokan. Security Associations in Personal Networks: A Comparative Analysis. In *Security and Privacy in Ad-hoc and Sensor Networks Workshop (ESAS)*, pages 43–57, 2007.
38. Ersin Uzun, Kristiina Karvonen, and N. Asokan. Usability analysis of secure pairing methods. In *Financial Cryptography and Data Security & International Workshop on Usable Security (USEC)*, 2007.
39. A. Varshavsky et al. Amigo: Proximity-Based Authentication of Mobile Devices. In *Ubiquitous Computing (UbiComp'07)*, pages 253–270, 2007.
40. Srdjan Čapkun and Mario Čagalj. Integrity regions: authentication through presence in wireless networks. In *WiSe '06: Proceedings of the 5th ACM workshop on Wireless security*, 2006.