Is that you, Alice? A Usability Study of the Authentication Ceremony of Secure Messaging Applications

Elham Vaziripour, Justin Wu, Mark O'Neill, Ray Clinton, Jordan Whitehead,
Scott Heidbrink, Kent Seamons, Daniel Zappala
Brigham Young University
{*elhamvaziripour,justinwu,mto,rclinton,jaw,sheidbri*}*@byu.edu*, {*seamons,zappala*}*@cs.byu.edu*

ABSTRACT

The effective security provided by secure messaging applications depends heavily on users completing an authentication ceremony—a sequence of manual operations enabling users to verify they are indeed communicating with one another. Unfortunately, evidence to date suggests users are unable to do this. Accordingly, we study in detail how well users can locate and complete the authentication ceremony when they are aware of the need for authentication. We execute a two-phase study involving 36 pairs of participants, using three popular messaging applications with support for secure messaging functionality: WhatsApp, Viber, and Facebook Messenger. The first phase included instruction about potential threats, while the second phase also included instructions about the importance of the authentication ceremony. We find that, across the three apps, the average success rates of finding and completing the authentication ceremony increases from 14% to 79% from the first to second phase, with second-phase success rates as high as 96% for Viber. However, the time required to find and complete the ceremony is undesirably long from a usability standpoint, and our data is inconclusive on whether users make the connection between this ceremony and the security guarantees it brings. We discuss in detail the success rates, task timings, and user feedback for each application, as well as common mistakes and user grievances. We conclude by exploring user threat models, finding significant gaps in user awareness and understanding.

1. INTRODUCTION

Recent disclosures of government surveillance and fears over cybersecurity attacks have increased public interest in secure and private communication. As a result, numerous secure messaging applications have been developed, including Signal, WhatsApp, and Viber, which provide end-to-end encryption of personal messages [19].

Most popular secure messaging applications are usable because they hide many of the details of how encryption is provided. Indeed, people are primarily using these applica-tions due to peer influence, not due to concern over privacy or security [5].

The strength of the security properties of these applications rests on the *authentication ceremony*, in which users validate the encryption keys being used. Unfortunately, there is evidence that most users do not know how to successfully complete this ceremony and are thus vulnerable to potential attacks [15]. Any user who does not execute the authentication ceremony for a particular conversation is essentially trusting the application's servers to correctly distribute the encryption keys. This leaves users vulnerable to threats that can intercept communications.

Several recent papers have shown that the authentication ceremony in secure messaging applications is difficult to use and prone to failure. A study of Signal showed that users, all of whom were computer science students, were highly vulnerable to active attacks [15]. A comparison of WhatsApp, Viber, Telegram, and Signal, found that most users were unable to properly authenticate [8], though after being instructed on what to do most users were subsequently able to authenticate after a key reset.

This state of affairs motivates our study, which examines to what extent users can successfully locate and complete the authentication ceremony in secure messaging applications if they are aware of the need for authentication. To answer this question, we conduct a two-phase user study of WhatsApp [22], Facebook Messenger [7], and Viber [21]. We chose these applications because of their popularity and their different designs. The authentication ceremony in WhatsApp uses either a QR code or a numeric key representation that users can compare. Viber presents a numeric key representation and provides functionality for users to call each other within the ceremony to compare the key. Facebook Messenger provides a numeric representation of the keys for both users. In addition to these differences, WhatsApp and Viber offer only secure messaging, while Facebook Messenger offers both insecure and secure messaging. We are curious as to whether the inclusion of an insecure messaging interface hinders the ability of users to find and successfully use secure messaging and the authentication ceremony.

In the first phase of our study, we asked 12 pairs of participants to complete a scenario where one participant needed to send a credit card number to the other participant. They were both instructed to verify that they were truly communicating with their partner (authenticity) as well as to ensure that no other party could read their messages (confidential-

ity). Participants were told the application would help them accomplish these goals.

In the second phase of the study, we presented 24 pairs of participants with the same task and scenario provided in the first phase. However, unlike the first phase, participants first read through an additional set of instructional slides before beginning the task. These slides informed them about traffic interception, that secure messaging applications use a "key" to secure conversations, and that to be secure they needed to confirm that they saw the same "key" as their partner. Participants were not instructed on how to use the applications to compare keys, nor shown any screenshots of the authentication ceremony; they were only told that each application had some way of providing this functionality. For both study phases, the method used for authentication was left to their discretion.

Each phase was a within-subjects study, and all participants engaged with all three applications in each phase. Participants differed between the two phases, allowing us to capture between-subjects differences in instruction between the two phases. We measured success rates in completing the authentication ceremony, time to both find and complete the ceremony, and user feedback on the applications, which includes System Usability Scale (SUS) scores, ratings of favorite application, ratings of trustworthiness for each application, and qualitative feedback.

Our findings include:

- In the first phase, despite the instruction about potential threats, the overall success rate over all participants and all applications was 14%, and only two of the twelve pairs of participants successfully located and completed the authentication ceremony. All other pairs attempted to authenticate one another through video calls, asking questions that required special knowledge to answer, or other ad hoc methods.

- In the second phase, the overall success rate increased to 79% for location and completion of the authentication ceremony. The success rates for the three applications were: 96% for Viber, 79% for WhatsApp, and 63% for Facebook Messenger.

- Viber's higher success rate was statistically significant when compared to the other two applications. This is interesting because Viber's authentication ceremony uses an in-app phone call and provides a UI that helps users view and read the encryption key during the phone call. Both WhatsApp and Facebook Messenger also provide manual verification of the encryption key, but do not provide this assistance. For both of these applications, numerous participants sent the key through in-app text, voice, and video, with a minority comparing the keys in person. Nearly half of participants chose to use the option WhatsApp provided for scanning a QR code.

- Averaged across the three applications, discovery of ceremony functionality took 3.2 minutes with ceremony completion necessitating another 7.8 minutes.

- All applications were rated in the "C" range on the System Usability Scale, indicating a need for significant usability enhancements.

- Most participants had not heard of Viber prior to their participation in our study. Trust ratings were very low in the first phase, but increased significantly in the second phase, when some instruction about security was received. This provides some evidence that learning about security features can enhance trust in a secure messaging application.

- Numerous participants complained about the length of the encryption key when having to compare it manually, taking shortcuts and often feeling fatigued by the process.

- Our qualitative data indicates that our participants have a healthy wariness for, and high-level understanding of: impersonation attacks, government and developer backdoors, and physical theft. They are, however, generally unaware of the existence of man-in-the-middle attacks, both passive and active. Our data is inconclusive on whether users make the connection between this ceremony and its security guarantees.

Our main takeaway is that even with an awareness of potential threats, users are not aware of and do not easily find the authentication ceremony in the secure messaging applications we tested. If given some instruction on the importance of comparing keys, they can find and use the authentication ceremony, and Viber's second-phase success rate indicates that a high success rate is a realizable goal. However, for all applications, the time to find and use the authentication ceremony is unsatisfactory from a usability standpoint. The natural tendency of our participants to use personal characteristics for authentication, such as a person's voice, face, or shared knowledge, indicates that future work could leverage this for a more user-understandable method of authentication.

2. RELATED WORK
Several papers have studied the usability of the authentication ceremony in secure messaging applications.

Two papers study the usability of the ceremony in a particular application. Schröder et al. studied Signal, showing that users were vulnerable to active attacks due to usability problems and incomplete mental models of public key cryptography [15]. This study included 28 computer scientists; of the participants, four clicked through the warning message, eight could not find the ceremony, and ultimately only seven were able to successfully authenticate their peer. Assal et al. asked participants to perform the authentication ceremony in ChatSecure using different key representations, which include a fingerprint, shared secret, and QR code [1]. Of the 20 participants in this study, 20% were successful for the fingerprint, 85% for the shared secret, and 30% for the QR code.

Two papers have compared the usability of various fingerprint representations. Tan et al. compared eight representations, including textual and graphical representations with varying degrees of structure, in a simulated attack scenario [18]. Graphical representations were relatively more susceptible to attack, but were easy to use and comparison was fast. Participants used different strategies for comparison, often comparing only a portion of the fingerprint or comparing multiple blocks at a time. Dechand et al. studied textual key

verification methods, finding that users are more resistant to attacks when using sentence-based encoding as compared to hexadecimal, alphanumeric, or numeric representations [6]. Sentence-based encoding rated high on usability but low on trustworthiness.

Herzberg and Leibowitz examined the usability of Whats-App, Viber, Telegram, and Signal, finding that most users were unable to properly authenticate, both in an initial authentication ceremony and after a key reset [8]. The study included 39 participants from a variety of backgrounds and all were given instruction on end-to-end encryption. Most users failed to authenticate on the first attempt; they were then given additional instruction about authentication. About three-quarters authenticated properly after the additional instruction was given.

Our work differs from these studies in several important ways. First, we study in detail the ability of users to discover and use the authentication ceremony in a variety of secure messaging applications, giving us insight into the differences among these applications. Schröder et al. only study Signal, and Dechand et al. do not study any particular applications. Second, we use a paired participant methodology, so that users are asked to identify a friend they already know, rather than an unknown study coordinator. This method is more realistic than most prior studies and yields important insights into user behavior. For example, our study participants called each other, verified through voice and vision, and asked questions based on shared knowledge. Third, we conduct a between-subjects study on the effects of instruction, so that those receiving instruction are not biased by their previous experiences. The first set of participants were asked to authenticate given only general awareness of threats, while the second set of participants received instruction about the importance of comparing encryption keys.

Another important aspect of our work is that it provides replicability that is not possible with prior work. Herzberg and Leibowitz report a similar result, that participants authenticated properly after additional instruction about authentication was given. However, their paper provides few details about the instruction given and does not report detailed statistics, so it is difficult to draw any quantitative conclusions about the effect of the instruction or the relative merits of the different applications they tested. We report detailed statistics about what methods users tried with each application, the time taken to authenticate, SUS scores, trust ratings, and favorite systems. We include our full study materials in the appendix and provide our dataset on a companion web site.

Significant work in the area of secure email has also examined issues related to usable authentication. Obtaining and verifying the key for a recipient is an important use case for email, and lessons learned may apply to secure messaging as well. Numerous papers attest to the difficulties users have with this and other key management steps [23, 16, 12].

The most success in this area has been in the use of automatic authentication using a trusted key server. Bai et al. [3] has shown that individuals recognize the security benefits of manual key exchange, but prefer a centralized key server that authenticates users and distributes keys associated with their email address, due to greater usability and "good enough"

security. This model has been simulated by Atwater et al. [2] and implemented using IBE by Ruoti et al. [11]. Likewise, the use of secure messaging applications is generally considered a success for automatic key management.

3. APPLICATION DESCRIPTIONS

The three secure messaging applications used in our study are WhatsApp, Viber, and Facebook Messenger. These three applications were chosen because they present users with distinct key verification experiences and because of their popularity and large installation base.

3.1 WhatsApp

WhatsApp is perhaps the most well-known and widely-used messaging application, boasting a user base of over one billion users. While it did not originally offer secure messaging functionality at its inception, in November of 2014, Whats-App partnered with Open Whisper Systems to incorporate end-to-end encryption using the Signal encryption protocol.

When a conversation is initiated, WhatsApp inserts a message informing users that messages they send are encrypted with end-to-end encryption. Users are given two options for key verification: QR code scanning and key fingerprint verification (both parties see the same fingerprint). In accessing this dialog, a short caption accompanies the "Encryption" option in the previous menu, informing users that they can "Tap to verify." Doing so brings up the verification dialog shown in Figure 1a.

3.2 Viber

Viber is another widely-used messaging application with an install base of over 800 million users. As with WhatsApp, it did not originally offer end-to-end encryption, adding this functionality in April of 2016. Its encryption protocol is a proprietary design allegedly based on the principles of the Signal protocol.

While—as with the other two applications—Viber does not immediately make apparent the need to verify keys, once begun, it does—unlike the other two applications—carefully guide the user through the process with a set of instructional dialogs. In displaying this functionality, Viber does not opt to use the terms "encryption" or "key" at the outset, instead characterizing the verification process as "trust[ing]" one's conversation partner. Only after the user selects this option, are they prompted with a dialog that explains the need to confirm that "secret keys are identical." This process is facilitated via a free Viber call. After making the call, both sides may see their keys by tapping a lock icon that appears during the call, allowing for verification. This dialog is shown in Figure 1b. It should be noted, however, that Viber does not allow the user to view their keys without initiating this call, nor does it allow the user to view these keys once a contact has been marked as trusted.
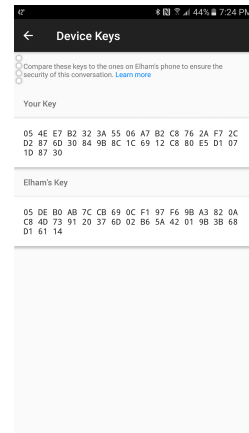
3.3 Facebook Messenger

Facebook Messenger is the messaging utility designed by Facebook to integrate into their chat system, and, like WhatsApp, has a user base of over 1 billion users. Again, as with the other two applications, it did not originally offer end-to-end encryption, adding this functionality in October of 2016. It also uses the Signal protocol.
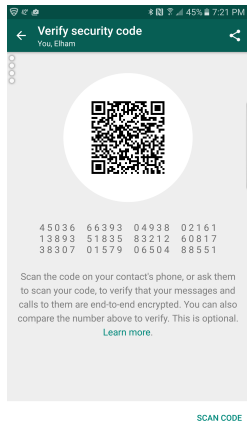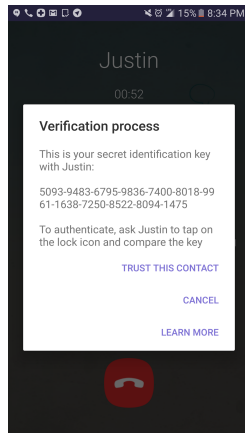
| (a) WhatsApp | (b) Viber | (c) Facebook Messenger |

Figure 1: Authentication ceremonies in each of the applications.

The user experience of Facebook Messenger's encryption functionality differs substantially from WhatsApp and Viber. While the first two applications encrypt all communication automatically, Facebook Messenger defaults to an unencrypted chat session, with users required to initiate a standard chat session before accessing a "Secret Conversation" function via the conversation menu. Once within the secret conversation menu, users can access their device keys via the context menu. At this point, the experience again diverges from the two other applications, as the key verification dialog presents users with two keys instead of one. Furthermore, the Facebook Messenger key verification interface does not easily facilitate a way for users to communicate these key values to the other party. This dialog is shown in Figure 1c.

4. METHODOLOGY

We conducted an IRB-approved, two-phase user study examining how participant pairs locate and complete the authentication ceremony in three secure messaging applications: WhatsApp, Viber, and Facebook Messenger. Our study materials are shown in Appendix B and our full data set is available at `https://alice.internet.byu.edu`.

In both phases, we asked participants to complete a scenario where one participant needed to send a credit card number to the other participant. We instructed participants to verify that they were truly communicating with their partner and to ensure that no other party could read their messages. Our instructions informed participants that the application would help them accomplish these goals, but they were left in control of the methods used to ensure these conditions were met. In the second phase, participants viewed and read aloud an instructional set of slides that informed them about the importance of comparing encryption keys.

Each phase was a within-subjects study, and all participants used all three applications in each phase. The participants differed between the two phases, allowing us to see between-subjects differences in instruction between the two phases.

To choose the three applications we compared the authentication ceremony in 10 secure messaging applications—WhatsApp, Telegram, Signal, Zendo, Facebook Messenger, Viber, Chat-Secure, Allo, Line, SafeSlinger. We binned the applications into groups, based on the authentication methods used. We then narrowed our choices to the following: Signal/WhatsApp (use both QR codes and manual verification), Telegram/Facebook Messenger (use manual verification, include non-secure chatting), and Zendo (uses NFC or QR code, requires verification before chatting). We chose WhatsApp over Signal and Facebook Messenger over Telegram because of their greater popularity in the United States. As explained below, we were unable to proceed with Zendo in the study. We chose Viber as an alternative because it provides a method for manually comparing encryption keys using a phone call built into the application. This provided us with three different applications that use a variety of authentication methods.

4.1 Pilot study

We conducted a pilot study of the first phase with three pairs of participants, using WhatsApp, Facebook Messenger, and Zendo. The Zendo secure messenger employs key verification as a forcing function: users must first scan each other's QR codes, or use NFC communication, before the conversation can begin. Unfortunately, we experienced multiple, severe technical difficulties with the application during the pilot study, leading us to abandon it in favor of Viber.

4.2 Study recruitment and design

We placed flyers advertising the study around the campus of a local university. These flyers contained a link that participants could use to schedule online, and they included a requirement that all participants bring a friend and smartphones in order to take part in the study. Recruitment proceeded from February 3, 2017 to February 28, 2017, with 39 unique participant pairs being recruited in total: 12 for the first phase of the study, and 24 for the second.[1]

---

[1]One second-phase participant pair experienced difficulty because one participant had limited English proficiency and our study was executed entirely in English (this participant thought that they were being tasked with locating a physical key). Technical errors occurred during the data collection of two other pairs and they were presented with incorrect post-task questionnaires. Accordingly, the data for these three pairs were excluded from the study and we recruited replacements in their place.
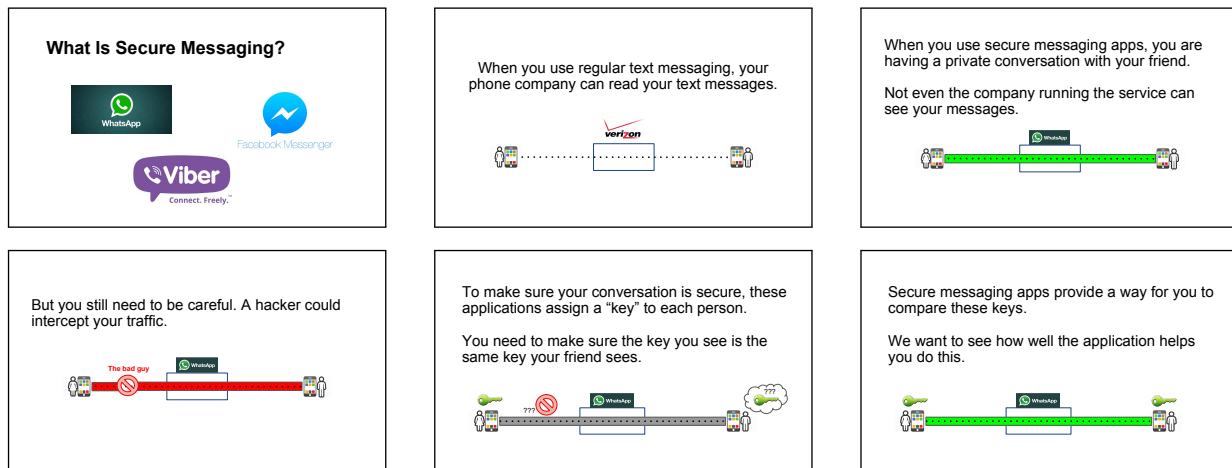
Figure 2: Instructional slides used in the second phase.

To ensure different pairs of participants tried applications in different orders, we calculated a complete set of permutations listing the order in which each of the three applications would be used by a given pair. We then randomized the permutation that was assigned to each participant. This ensured a collectively uniform distribution of sequences while keeping the assignment of a given sequence to a particular pair random. Each ordering of the three systems occurred exactly twice in the first phase and four times in the second.

The study was conducted in two phases, spanning a period of one month. The first phase ran from February 3, 2017 to February 16, 2017. It took roughly 40 to 45 minutes for each pair of participants to complete, for which they were compensated $10 USD each. The second phase ran from February 17, 2017 to March 2, 2017. The second phase studies were more involved and took longer to complete, roughly 60 minutes each, and so all participants were compensated at a higher rate of $15 USD.

When participants arrived for their scheduled appointment, we presented them with the requisite forms for consent and compensation. We instructed them to download and install any of the three applications—WhatsApp, Viber, and Facebook Messenger—that they did not already have on their phones, to minimize the likelihood of technical difficulties during the study.[2] We then read them a brief introduction describing the study conditions and their rights as study participants. We informed them that they would be placed in separate rooms, but could freely communicate or meet with one another if they deemed it necessary to complete their task. We also informed participants that a study coordinator would be with them at all times and would answer any questions they might have.

We randomly assigned one member of each pair as Participant A, with his or her counterpart becoming Participant B, delineating their roles in the subsequent tasks. We then led them to their respective rooms, seating them at a computer, and initiating audio recording. We preloaded each computer with a Qualtrics survey that guided participants through the study, and it included a demographic questionnaire, instructions regarding the three tasks they were to perform, and post-task questionnaires. Each of the three tasks was identical in nature, differing only by which of the three secure messaging applications participants were to use to complete the task. Throughout the study, study coordinators were available to answer general questions about the study, but were careful not to provide any specific instructions that would aid in the use of the applications themselves.

4.3 Task design

In both phases, the tasks participants completed were the same: Participant A was to securely retrieve a credit card number in Participant B's possession by using the application that was being tested. This scenario was intended solely as a narrative backdrop for the tasks we were truly concerned with: finding and completing the authentication ceremony. When asked to complete the task, participants were instructed as follows:

> Your task is to make sure that you are really talking to your friend and that nobody else (such as the service provider) can read your text messages. The application should have ways to help you do this.

Accordingly, despite a difference in roles, there were no practical differences between the tasks Participant A and Participant B needed to complete. Participants were instructed and encouraged to "talk aloud" as they completed the task, explaining the choices they made and the actions they took.

Additional instruction was given in the second phase. Before participants were introduced to the task, they were asked to read aloud a short set of slides, shown in Figure 2. These slides informed them that traffic interception was a possibility, that secure messaging applications accordingly provide a "key" that could be compared to ensure that conversations were indeed secure, and that they needed to make sure that they saw the same key as their counterpart. Furthermore, on the

---

[2]In our pilot study, several participants lacked sufficient space on their phones to install the applications or had phones that were too old to run the applications properly. We subsequently adopted this measure in an attempt to forestall re-occurrence.

first task, if second phase participants had failed to verify one another's identity either prior to sensitive data exchange or after ten minutes had passed, they were marked as having failed the task and prompted by study coordinators to look for a way to authenticate properly.

4.4 Study questionnaire

Participants were led through the study by a web-based Qualtrics survey. We first discuss those aspects that were held constant for both phases, followed by an explanation of how the questionnaire differed in the second phase.

Upon beginning the survey, participants first answered a set of demographic questions. They then answered questions about their past experience, if any, with secure messaging applications. This included questions about which applications they might have used, their reasons for doing so, and their general experiences with sending sensitive information. Participants were next shown a description of their first task (all three tasks were identical, diverging only on the system being used). Each task was followed with a post-task questionnaire assessing their level of trust in the application, whether or not they believed they had successfully verified their partner's identity and why, and who they believed was capable of reading their conversation. After all three tasks had been completed, participants were then asked which of the three applications was their favorite and why.

In the second phase, participants were given supplementary instructions and asked additional questions. First, after the demographic questions, participants were asked a series of six questions intended to gauge their relative familiarity with end-to-end encryption. Next, prior to beginning the first task, they were presented with, and asked to read aloud, a set of six slides that very briefly introduced the role of keys and informed them that the applications they were about to use would provide a way for them to compare these keys. These instructional slides are shown in the appendix. Finally, at the end of each task, the post-task questionnaire from the first phase was augmented by the ten questions from the System Usability Scale (SUS).

4.5 Post-study debrief

At the conclusion of each study, participant pairs were gathered in the same room and asked a series of questions. This served as a complement to the questionnaires that they had answered individually, and gave them an opportunity to react to one another. Participants were prompted regarding incidents specific to their experience—e.g., if they had evidenced visible frustration with a particular app—as well as general questions. Examples of the latter include having participants contrast the authentication ceremony used by each application, as well as asking them to explain what role they thought keys played in verifying one another's identity.

4.6 Demographics

Our sample population skewed slightly female (n=40, 56%) and young, with 74% (n=53) between the ages of 18 and 24, and 26% (n=19) between 25 and 34. Because we distributed recruitment flyers on a university campus, most of our participants were college students (n=48, 61%), with 17% (n=12) having less educational experience than that, and 22% (n=16) having at least finished college. Participants had a variety of backgrounds, with roughly even representation between technical (i.e., STEM; n=34, 48%) and non-technical backgrounds (n=37, 52%), and 10 (14%) in explicitly IT-related fields. (One participant failed to identify their field of study or occupation.)

In the second phase, the questionnaire included a series of six multiple-choice questions intended to assess participants' knowledge of end-to-end encryption. We assigned equal weights of one point to each question, and scored each participant from 0-6, corresponding to the number of correct answers given by the participant. Participants were further placed into categories of "beginner," "intermediate," and "advanced" for scores in the range of 0-2 for beginners, 3-4 for intermediate, and 5-6 for advanced. There were an equal number of participants with beginner and intermediate ratings—21—with 6 participants netting an advanced rating. Beginners were mostly female (3:18), intermediate participants were mostly male (15:6), while the advanced category had an even gender split (3:3).

4.7 Limitations

The instructions given to the first three participant pairs of the first phase were slightly different from those given to the remaining nine. They were directed to ensure that no one was "listening in" on their conversation, a directive participants took literally as they would visibly scan the room for potentially intrusive parties. This wording was subsequently altered, with participants instead instructed to ensure that "nobody else (such as the service provider) can read your text messages."

The slides we provided participants to teach about cryptographic keys were necessarily simplified so that they could be understood by novices. In this material we mentioned that participants should ensure the key they see is the same as their partner's. While this was sufficient in describing tasks for Viber and WhatsApp, Facebook Messenger actually utilizes two keys, one for each partner. This subtlety was not mentioned by any participant nor did it seem to adversely affect their performance.

Finally, due to our method of recruitment, our participants were largely students and their acquaintances, and subsequently exhibited some degree of homogeneity, e.g., all participants were between 18 and 34 years of age. They are thus not representative of a larger population. Furthermore, while an effort was made to place participants in a more organic setting—e.g., by having them communicate with real members of their social circle as opposed to study coordinators—this was still ultimately a lab study and has limitations common to all studies run in a trusted environment [10, 17].

5. FIRST PHASE RESULTS

In the first phase of the study, only 2 of the 12 pairs experienced some success in locating and completing the authentication ceremony, with an overall success rate of 14% across all pairs and applications.

Participants used a variety of ad hoc methods for authentication. Listed in the order they appear in Table 1, these methods were: utilization of a picture for visual identification, utilization of a live video feed for visual identification, utilization of shared secrets for identification, utilization of contact information (e.g., phone number, profile picture) for identification, utilization of a shared second language for

Application	Send Picture	Recognize Video	Recognize Voice	Shared Knowledge	Contact Info	Second Language	Authentication Ceremony
WhatsApp	0	0	13	10	3	2	2
Viber	0	10	4	7	2	2	4
Facebook Messenger	2	12	2	7	0	0	2

Table 1: Methods of authentication used in the first phase by pairs of participants.

identification, and performing the actual authentication ceremony. These categories were compiled by asking users how they authenticated the other party, and are not mutually exclusive (some used more than one method).

We examined the two pairs that were successful to better understand their experiences. One pair was successful because of their curiosity, which led to them exploring the application settings. This pair started with Viber and began to verify each other simply through a phone call, when they suddenly noticed the option in Viber to authenticate a contact, making that contact "trusted." They subsequently verified the encryption key through the phone feature embedded in the authentication ceremony. After this experience, this pair noticed they should be looking for similar functionality in the other applications. The followed the on-screen instructions in WhatsApp to scan the QR code, and they exchanged a screenshot of the authentication code in Facebook Messenger.

A second pair started the study with Facebook Messenger. This pair called each other using an insecure phone call, spoke in Korean, and transferred the credit card number used in the scenario without completing the authentication ceremony. They next used WhatsApp, and because it was their first time using the application, they were prompted with a notice about end-to-end encryption after sending their first message. After clicking to learn more, this pair was able to locate and complete the authentication ceremony by using a phone call to read and verify the key. After this experience, the pair was also able to locate the lock icon in Viber, follow the instructions in the ceremony, and use a phone call to verify the key. However, they were unsure about the role of the key and still verified each others' identity by asking questions that relied on their common knowledge.

6. SECOND PHASE RESULTS
In this section we discuss results regarding participant use of the authentication ceremony for the second phase, when additional instruction was given regarding the importance of comparing keys.

6.1 Success Rate
The success rate for completing the authentication ceremony in the second phase was drastically higher than for the first phase. Overall, the success rate was 78% across all participant pairs and the three applications. Table 2 shows the breakdown of the success rate for each application. Failures occurred when participants transmitted sensitive data before verifying keys, or if they failed to find and validate the keys within ten minutes of opening the application. Successes indicate that participants identified and compared keys in some fashion. The Error column indicates three cases where Facebook Messenger failed to deliver messages or failed to display important UI elements that allow participants to access key information. We noted various mistakes made by

Application	Success	Fail	Error
WhatsApp	19 (79%)	5 (20%)	0 (0%)
Viber	23 (96%)	1 (4%)	0 (0%)
Facebook Messenger	15 (63%)	6 (25%)	3 (13%)

Table 2: Success rates per pair of participants for the authentication ceremony in the second phase.

participants, but these were considered distinct from failures and are discussed later.

The leap from a 14% success rate in the first phase to 78% in the second phase suggests that users are capable of locating and performing the authentication ceremony when prompted. Some of these applications indicate that keys need to be validated, yet our results from phase one indicate that these instructions are largely ignored, thus we suspect that the independent prompts from our study accounted for much of the difference seen in authentication ceremony success rates.

To test whether there are any differences between the applications, we used Cochran's Q test. We found that the success rate was statistically different for the applications ($\chi^2(2) = 15.429$, $p < .0005$). We then ran McNemar's test to identify the significant differences among the pairs of applications. We found there is a significant difference between WhatsApp and Viber ($p = 0.008$) as well as between Facebook Messenger and Viber ($p < 0.0005$).

It is interesting that Viber's success rate is significantly higher than the other two applications. Viber's authentication ceremony uses an in-app phone call and provides a UI that helps users view and read the encryption key during the phone call. Both Facebook Messenger's authentication also provides only manual verification of the encryption key, but does not provide this assistance.

6.2 Verification Methods
The methods used by participants to perform the authentication ceremony are shown in Table 3. Note that some participants used more than one method. We do not include methods for three pairs of participants who encountered errors when utilizing Facebook Messenger. These errors prohibited us from assessing how these participants would have interacted with the authentication ceremony.

The most-selected method for the ceremony through WhatsApp was scanning the QR code of the key fingerprint in person. Of the applications we studied, this method is unique to WhatsApp. Some pairs opted to take a screenshot of the key or QR code and send it this way, while others remembered substrings of the key fingerprint and repeatedly visited the text screen to send pieces of it to their partner. This behavior

Action	WhatsApp	Viber	Messenger
Secure Methods			
Scanned QR code in person	11 (46%)	N/A	N/A
Read key in person	1 (4%)	0 (0%)	7 (29%)
Called out of band or used Viber's call method to provide key	1 (4%)	23 (96%)	1 (4%)
Less Secure Methods			
Sent key through in-app text	7 (29%)	N/A	10 (42%)
Sent key through in-app video	3 (13%)	N/A	4 (17%)
Sent key through in-app voice	1 (4%)	N/A	1 (4%)
Failures			
Sent sensitive information before validation	5 (21%)	1 (4%)	5 (21%)
Failed to find key within 10 minutes and after a hint	1 (4%)	0 (0%)	1 (4%)

Table 3: Methods used for the authentication ceremony in the second phase. Numbers indicate pairs and percentages are out of the total number of pairs.

occurred when participants discovered the QR code and key fingerprint but were confused as to what to do next.

Numerous participants using WhatsApp read the key data in person, read the key using a voice or video call, or sent the key using text. Most participants using Facebook Messenger used these methods, since they were the only ones available.

Viber provides a much stricter interface once a user has located the option to verify his partner's identity. Instead of offering key material immediately, an in-app call must be initiated before the key material is provided to the user. As a result, all pairs who successfully completed the ceremony utilized this feature to verify their keys. We note that this policy resulted in no mistakes made for the authentication ceremony. However, the process confused some participants, and three pairs sent sensitive information through the application without performing this procedure.

6.3 Timing
We timed each pair of participants to obtain two metrics: the time taken to locate and identify the authentication ceremony as it is presented within the application interface and the time taken to complete the ceremony successfully. In the case of finding the ceremony, the time reported is the time taken for the first partner to identify the key material or complete the task. We consider timing data only for cases where the pair succeeded in authenticating successfully because we stopped participants after 10 minutes if they could not find the ceremony.
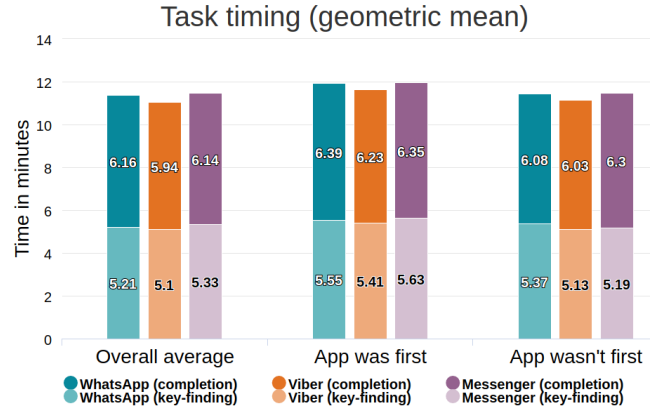

Figure 3: Timing for finding and using the authentication ceremony in the second phase. Lighter shades indicate the time taken to find the ceremony and the full bar indicates time taken for completing the ceremony.

Figure 3 shows the geometric mean of both time metrics for the three applications tested.[3] Applications that are selected to be evaluated first in a given study have a disadvantage with respect to time because it is users' first exposure to the task and possibly keys in general. To account for this, Figure 3 also includes comparisons showing timing data from when each application was studied first and when the application was not studied first.

To test whether there is a significant difference in the time to complete these tasks among the three different applications, we used the Kruskal-Wallis test. We found that there are statistically significant differences among the applications for completing the ceremony ($p = 0.043$). We next ran pairwise post-hoc Dunn's tests to determine where the differences occur. We found a significant difference between Viber and WhatsApp for completing the ceremony ($p = 0.045$), with Viber being faster (mean time, Viber=6.9 minutes, WhatsApp=8.5 minutes).

A major takeaway from the timing data shown is that key discovery and key verification both require substantial time for all three applications. On average, across all applications discovery of the ceremony required 3.2 minutes and ceremony completion required another 7.8 minutes. Given that the participants were informed about the existence of the keys beforehand and told explicitly to verify them, these times are unsatisfactory from a usability standpoint. The usability issues and concerns voiced by participants responsible for these times are discussed in Section 7.

7. APPLICATION FEEDBACK
In this section we discuss feedback that participants provided regarding the secure messaging applications, including usability, their favorite application, and the trustworthiness of the applications.

7.1 Usability
During the second phase of our study, participants evaluated each application using the System Usability Scale (SUS).

---

[3]Sauro and Lewis recommend using the geometric mean for task timing [14] because timing data is positively skewed and the geometric mean reduces error.

SUS subcategory	WhatsApp	Viber	Messenger
Overall	65.45	67.45	**67.78**
First system	65.47	67.97	**69.22**
Not first system	64.45	66.02	**67.97**
Success	64.41	67.86	**72.71**
Failure	66.25	63.13	**69.50**

Table 4: SUS scores for the applications in the second phase.

Study phase	WhatsApp	Viber	Messenger	None
One	39.1%	8.7%	47.8%	4.4%
Two	31.3%	22.9%	43.8%	2.0%

Table 5: Participants' favorite applications. Each cell contains the fraction of participants from each phase who, when prompted for their favorite system, gave the respective response.

Table 4 presents the breakdown of the scores for each system across various subcategories. The values shown are the mean values for each subcategory, while bolded values highlight the highest SUS score for each subcategory.

We report SUS scores across five subcategories for each application: overall SUS score, the mean SUS score when the application was the first of the three presented, the mean SUS score when the application was not the first shown, the mean SUS score for participants who succeeded at the task using the given application, and the mean SUS score for participants who failed the task.

Although SUS scores range from 0 to 100, this is not a percentile value and can thus be difficult to interpret. Accordingly, to help contextualize the values shown, we draw on the findings of researchers familiar with SUS. Sauro [13], extending work from other researchers such as Bangor et al. [4], created a guide for interpreting a given SUS score by normalizing it relative to those achieved by other systems. This framework associates SUS scores with percentile rankings and with letter grades (from A+ to F).

For reference, the applications' overall SUS scores fall within the "C" range, landing somewhere within the 41st to 59th percentile. The single lowest SUS score—Viber's mean failure score—nets a "C-" grade, falling within the 35th to 40th percentile. The highest SUS score—Facebook Messenger's mean success score—achieves a "C+" grade, somewhere within the 60th to 64th percentile.

7.2 Favorite application

Participants were asked to select which, if any, of the three applications was their favorite and why. Table 5 shows the breakdown of responses for each phase. Facebook Messenger was the most preferred system, followed by WhatsApp. We ran a Chi-Square test to determine if the differences in the ratings between phase one and phase two were statistically significant and they were not.

Though numerous reasons were given for why a particular system was a participant's favorite, familiarity was by far the most commonly cited reason for preference (except with Viber, which was not previously used by any of our participants). The next most common reason given, and one that held true for each of the three systems, was ease-of-use, with what constituted "easy to use" varying from system to system. Some WhatsApp users, for example, appreciated its ability to scan QR codes for key verification, obviating the need to read aloud the long string of digits comprising a key fingerprint. Those who liked Viber found its key verification process the simplest to access and execute. By contrast, those who mentioned ease-of-use relative to Facebook Mes-
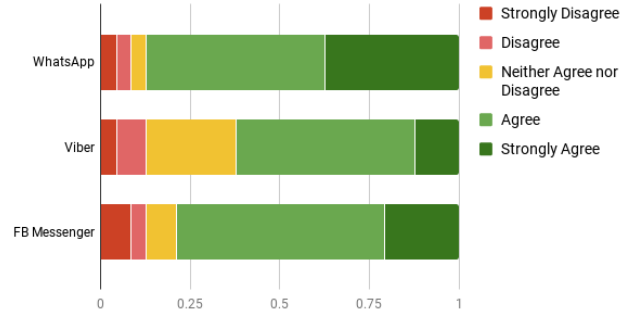
(a) Trust ratings in the first phase.
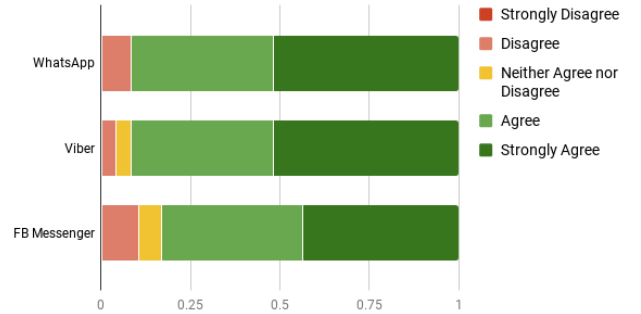
(b) Trust ratings in the second phase.

Figure 4: Participant ratings of trust for each application.

senger typically associated it with familiarity as opposed to any mechanism in particular.

7.3 Trust

As part of each post-task questionnaire, participants were asked to rate their trust in each application. They were presented with the statement *"I trust this application to be secure"* and asked to rate the statement on a 5-point Likert scale ranging from "strongly disagree" to "strongly agree." Responses for the two phases are shown, normalized, in Figure 4.

Comparing the trust scores from the two phases, two points stand out. First, a "strongly disagree" response—indicating a total lack of trust in the application—appeared for all three of the applications in the first phase, but not at all in the second phase. This is mostly due to one participant from the first phase who chose "strongly disagree" for all three systems. Secondly, responses of "strongly agree"—indicating confidence and trust in the application—are much more prevalent in the second phase.

To compare the trust scores in more detail, we ran a mixed model ANOVA Test, which allowed us to see the the interaction between the two independent variables (application and phase). We found that there is a significant interaction between the application and the study phase ($F(2,140) = 5.023$, $p = 0.008$, partial $\eta^2 = 0.067$).

To determine whether there was a simple main effect for the application, we ran a repeated measures ANOVA on each phase. There was a statistically significant effect of the application on trust for phase one ($F(2,46) = 4.173$, $p = 0.022$, partial $\eta^2 = .154$). By examining the pairwise comparisons, we found that the trust score was significantly lower for Viber as compared to WhatsApp in the first phase (M = 0.542, SE = 0.180, $p = 0.19$).

To determine whether there was a simple main effect for the study phase, we ran a one-way ANOVA on each application to compare the trust between the two phases. There was a statistically significant difference in trust ratings between the two phases for Viber ($F(1,70)=14.994$, $p < 0.0005$, partial $\eta^2 = .176$). The mean trust for Viber in the first phase was 3.58, and in the second phase it increased to 4.40.

Altogether, this analysis indicates that Viber was trusted less than WhatsApp in the first phase, but then was trusted significantly more in the second phase, after some instruction about the importance of the authentication ceremony. The trust for Viber increased in the second phase to the point that it was not significantly different from WhatsApp.

Participant commentary raised two other points of interest. First, participants strongly associated reputation with the trustworthiness of applications. Viber, for example, despite possessing a large user base outside of the United States, was essentially unknown to our participants, leading them to express wariness of this application. Facebook's status as a household name both inspired confidence and distrust. While its reputation as a large and established company reassured some, others were discomfited by the many negative stories they had heard about account hacks and privacy invasions on Facebook. Second, responding to descriptions of end-to-end encryption and promises of secure communication by the various applications, multiple participants remarked that they had no way to truly gauge the validity of those statements. Both these sentiments are captured by a remark from R10B, *"I would say it's a double-edged sword because Facebook—everyone knows Facebook—but it has that reputation of getting hacked all the time. But I've never heard of Viber or WhatsApp, so it could easily be some third-party Ukrainian mean people who want to steal information because that's just who they are. And whether it states that they're not gonna read or listen to the conversations and stuff like that... well, who knows?"* However, most opted to believe, for as one participant concluded, *"at some point, you have to trust* **something***."*

8. OBSERVATIONS
During our study, certain participant experiences and commentary stood out, highlighting a handful of concerns about each of the three applications individually, and in general. We feel that these observations are worthy of note in that they suggest directions for focus and improvement in the domain of secure messaging.

8.1 Single key fingerprint
WhatsApp and Viber both generate a single key fingerprint to be shared between pairs. While alternating recitation of segments of the key is likely the intention of developers, in practice, relationship dynamics complicate the issue. We observed several instances where the dominant partner in the relationship read the entire key on their own, with their partner simply offering an affirmation in response. When key verification is done in this manner, one party never actually demonstrates any knowledge of the shared secret—it is entirely possible that a man-in-the-middle could simply convey validation of the key when their key is, in actuality, different. This effect is further emphasized when, as we saw in one instance, the listening party asks the speaking party to repeat the first part of the key, reinforcing the speaking party's belief that their partner is in possession of the correct key. It is, however, worth noting that this "extended" validation once again did not demonstrate any actual knowledge of the secret.

8.2 Key length
It was often observed during the study that participants were surprised at the length of the key data they were intended to relay to their partners. Though every application used a form of fingerprinting to greatly reduce the total characters that needed to be read, users often verbally remarked that strings were too long. During the key exchange process we often witnessed fatigue, as participants would read only half the key and claim that was "good enough" and some recipients even ignored the key being read to them by their partners after the first few numbers matched. R27A used a QR code transmission to handle her first authentication ceremony with WhatsApp. Upon realizing that no such option existed for Viber, her second application used, she looked at the key and exclaimed, *"It's about eight years long!"*. R27A successfully checked every digit of the key data with her partner, but voiced her disapproval of its length repeatedly throughout.

8.3 Viber-specific issues
We observed two issues with Viber. The first relates to its mechanism for verifying a new user's phone number. While most applications send a confirmation text containing a code, as does Viber, it nevertheless defaults to calling the new user first as a primary and alternative confirmation mechanism. This took many of our participants by surprise and left them ill-at-ease to see an unknown number suddenly calling them. Secondly, and far more concerningly, Viber does not provide a mechanism to revoke trust. While this is likely a conscious decision on the developers' part, it can cause issues in practice. More specifically, one participant inadvertently tapped the trust button while trying to figure out how to verify his partner's key, thus accidentally conveying to the application in an apparently irreversible manner that this individual was now trusted.

Many users were also critical of the Viber UI's phrasing for the option to begin the process of key verification. The option is labeled "Trust this contact," which many users hesitated to press, unsure if it would inform the application to trust the contact or if it would bring up further dialogues to perform the validation. R36A visibly hesitated during this step during the study and articulated this concern in the exit interview: *"if I click 'Trust this Contact' but I haven't verified [my partner] yet, it's kind of weird."*

8.4 WhatsApp-specific issues

We observed several issues with WhatsApp. WhatsApp appends a pair of checkmarks next to each message, representing the delivery and read status of the respective message. However, a handful of participants mistakenly associated these checkmarks with security, operating under the misconception that a checkmark beside a given message indicated that it had been secured. The other two issues concern the key verification mechanism. When a matching QR code is scanned, the application briefly flashes a green checkmark logo over the QR code area, indicating that the fingerprint has been validated and is correct. However, because it disappears quickly, leaving no lasting indication that verification has occurred, numerous participants wondered if they had verified the key or not. Additionally, the key verification screen includes a button to share a screenshot of the verification screen. Some of our participants assumed that they could use this to send a screenshot to their partner, who could then scan the QR code contained therein. Unfortunately for them, WhatsApp does not provide functionality to scan a QR code from an image, serving to confuse those who tried.

8.5 Facebook Messenger-specific issues

In addition to the usability concerns already described, such as the difficulty in locating device keys, Facebook Messenger's Secret Conversation functionality—its mechanism for secure communication—errored more than a few times during our study. More importantly, however, was that these errors were not apparent to participants. Participants were thus unaware that the Secret Conversation was not operating as intended, and instead blamed themselves or their counterparts for failure. One example we encountered several times was that encrypted messages sent via this mechanism appeared normally on the user's phone despite never being received by their partner. One such participant began shouting in exasperation at her phone, exclaiming, *"I feel like I am having a conversation with myself! What's wrong with this app?!"*

8.6 Key changes

One important issue that secure messengers must deal with in practice is a key change occurring mid-conversation. As this was not tested by our participants during our study, we recreated this scenario in each of the three applications to observe their respective reactions. Facebook Messenger inlines a message when one's conversation partner's key changes, informing the user that their device has changed and that their key has changed. While it does not explicitly instruct the user to re-verify the key, of the three applications, it makes the user aware that key change has occurred. Viber gives no proactive notification to the user that key change has occurred, but when the conversation menu is again accessed post-change, Viber includes an explicit message warning the user that they will need to re-verify the identity of their conversation partner. WhatsApp presented no notification that we could observe. It neither inlined a notification as Facebook Messenger did, nor does it indicate to the user that re-verification must be performed. In fact, WhatsApp presents no lasting UI change that allows a user to confirm that verification has occurred at all.

9. USER THREAT MODEL

Two authors jointly coded responses to two survey questions used in both phases regarding participant perception of the authentication ceremony. These questions were:

- Please explain why you think you have (or have not) verified the identity of your friend.

- Who do you think can read your message except you and your friend?

In reviewing the coded data, some details of the threat models perceived by users became evident.

Note that, if correctly followed, completing the authentication ceremony successfully guarantees that a participant has authenticated their partner and no other party can listen in on the conversation. This of course assumes that the applications have properly implemented cryptographic protocols. None of the applications studied are open source, so their claims cannot be verified.

Of the 141 times the first of these prompts was presented (excluding Facebook Messenger errors), 109 responses indicated that the authentication ceremony was a primary reason for successful identification. This is encouraging, but also expected given the focus that the study placed on its significance, which may have biased participants. For example, in response to the first prompt, R13B stated *"...I asked him a person[al question] that he responded [to] in the right manner, but also because our messages were encrypted and our personal keys matched."* The use of questions that rely on shared knowledge was a common response to this prompt, and it was often coupled with a reference to verifying the key.

Where verification of personal inquiries are mentioned in tandem with key verification as a reason for verified identities, it is unclear whether participants believe the inquiry can be used as a substitute for key verification or if they are expressing the more secure notion that proper key verification includes explicit identity matching. To mitigate any mislabeling due to this lack of clarity, we focus on the responses that did not mention key verification as the reason for identity verification, which occurred 32 times. These responses focused on verifying features of their partner and considered impersonation or physical duress attack vectors. For example, R24A asserted he had verified the identity of his partner because he had *"asked personal questions that are difficult to know from online material/searches* and R36B confided that his partner *"was able to tell [him] something that no one else would know. Unless he was being held at gunpoint."* Of these 32 responses, 28 (88%) of them mention using features of their partner as the method of verifying identity (e.g. physical appearance in video, shared private knowledge, familiar voice). Two others mentioned trust in the application itself, one admitted no attempt to verify, and one trusted that their partner verified on their behalf.

The second prompt listed above provided some insight into the set of possible attackers considered by participants. This question was issued 141 times as well, immediately following the prompt mentioned earlier. Though 109 responses indicated that the identity of their partner had been verified, only 76 (70%) responses indicated that no other party could read messages exchanged between the two partners. The responses of those who indicated that other parties may be privy to the information were coded to determine the nature of the suspect parties. Five distinct entities were found to be mentioned in those responses: government, cellular service providers, physical accessors (e.g., shoulder surfers,

Type	Times Mentioned
Service Provider	4
Government	8
Hackers	17
Physical Accessors	18
Application Developer	19

Table 6: Attacker types suspected by participants.

thieves), the application developer, and remote "hackers". The number of times each of these entities was mentioned in a response are recorded in Table 6. Thirty-three of these labels come from persons who identified the importance of the key in verifying their partner's identity but obviously remained skeptical as to the full security of the application.

It is interesting to note that man-in-the-middle attacks were not explicitly mentioned as a possible attack vector in the responses to either of the prompts evaluated here. Impersonation was mentioned frequently in responses to the first prompt, and various tampering by governments and those with physical access to phones and their software were mentioned in responses to the second prompt. The apparent lack of awareness of man-in-the-middle attacks seemed to influence the trust users had in each other's identity, based on the frequent mentions of things like shared knowledge and videos used when identifying users. Many respondents further demonstrated this unknown attack surface through additional commentary. For example, R24A said he *"just did not consider verifying her identity. Thought [it] would [be] hard to replicate it within this short time."*

Many users did seem to grasp that there were other attacks possible, but used the term "hacker" as a generic catchall for these. For example, R27B mentioned that no one could read the messages sent between her and her partner *"unless people read over our shoulder or people hack into our Facebook accounts and read them before we delete them."* Similarly, R36A and R28A stated that the only people who could read the encrypted messages were *"just the two of us unless there were hackers"* and *"not WhatsApp or third parties! But probably people with skills,"* respectively.

In addition to being a catchall, use of the "hacker" response may also be providing insight into belief in a theoretical ceiling of network security by users. Since most users are unfamiliar with the mathematical foundations of cryptography and the details of security protocols, many struggle to adopt secure practices and understand the nature of various threats. On the other hand, users are often aware of their own ignorance in such matters, and these responses might indicate that users account for this in mental models by incorporating a "hacker" entity that is always capable of subverting any piece of the system. In this sense, lack of security knowledge affects both users' ability to make secure decisions *and* lowers their confidence in security itself.

Some users also expressed some suspicion of the applications themselves for government and/or developer eavesdropping. R24B was suspicious of both: *"Viber (if they want to) & government investigation agencies".* Others respondents explicitly mentioned "backdoors" built into the applications or general suspicions like R29B: *"I still feel like WhatsApp can*

read the messages even though they say they can't." Finally, some users were wary of logging, as exemplified by R15A: *"The company I'm sure has records of the texts but [security] depends on if they go through them or not."*

Overall, the responses indicate that users have a healthy wariness and high-level understanding of impersonation attacks, government and developer backdoors, and physical theft, but that the same cannot be said for man-in-the-middle attacks, both passive and active. It is assumed that some of the mentions of "hackers" refer to this, but these responses were far less specific than for other attacks. In other words, it appears that users' threat models do not include the ability for attackers to be positioned in between the two endpoints of a conversation. If this was understood, we hypothesize that far less respondents would have relied on physical appearance or shared knowledge as an identity verification mechanism. Since one of the primary goals of the secure exchange of keys is to thwart man-in-the-middle attacks, work may be needed to help users understand this attack vector.

10. CONCLUSION

We used a two-phase study to examine whether users are able to locate and complete the authentication ceremony in secure messaging applications. In the first phase, users were aware only of the need to authenticate and ensure confidentiality, and only two of twelve users were able to locate the authentication ceremony, with an overall success rate of 14%. Participants instead primarily used personal characteristics, such as a person's voice, face, or shared knowledge. In the second phase, users were instructed about the importance of comparing encryption keys in order to authenticate a partner, leading to an overall success rate of 78%. Users were significantly more successful using Viber. However, the time required to find and use the authentication ceremony was 11 minutes, combined, on average across all applications, which may be so long that it would discourage users from authenticating each other.

Based on our findings, we believe that many users can locate and complete the authentication ceremony in secure messaging applications if they know they are supposed to compare keys. However most people do not understand the threat model, so it is not clear that they will know how important it is to compare keys.

An open question is how secure messaging applications can prompt the correct behavior, even without user understanding. It may be possible to leverage the tendency users have to rely on personal characteristics for authentication. We are exploring the use of social authentication [20] as a way of translating authentication of encryption keys into a method that is more understandable to users.

Another area for future work is improving the authentication ceremony so that it does not take so long to complete. A system like CONIKS [9] may help to automate the process of discovering another person's key without relying on a single trusted party, while also providing non-equivocation so that key servers cannot deceive users.

11. ACKNOWLEDGMENTS

12. REFERENCES

[1] H. Assal, S. Hurtado, A. Imran, and S. Chiasson. What's the deal with privacy apps?: A comprehensive exploration of user perception and usability. In *International Conference on Mobile and Ubiquitous Multimedia*. ACM, 2015.

[2] E. Atwater, C. Bocovich, U. Hengartner, E. Lank, and I. Goldberg. Leading Johnny to water: Designing for usability and trust. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 69–88, Montreal, Canada, 2015. USENIX Association.

[3] W. Bai, D. Kim, M. Namara, Y. Qian, P. G. Kelley, and M. L. Mazurek. An inconvenient trust: User attitudes toward security and usability tradeoffs for key-directory encryption systems. In *Twelfth Symposium On Usable Privacy and Security (SOUPS 2016)*. USENIX Association, 2016.

[4] A. Bangor, P. Kortum, and J. Miller. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies (JUS)*, 4(3):114–123, 2009.

[5] A. De Luca, S. Das, M. Ortlieb, I. Ion, and B. Laurie. Expert and non-expert attitudes towards (secure) instant messaging. In *Twelfth Symposium On Usable Privacy and Security (SOUPS 2016)*. USENIX Association, 2016.

[6] S. Dechand, D. Schürmann, T. IBR, K. Busse, Y. Acar, S. Fahl, and M. Smith. An empirical study of textual key-fingerprint representations. In *Twenty-Fifth USENIX Security Symposium (USENIX Security 2016)*. USENIX Association, 2016.

[7] Facebook. facebookmessenger.com. `https://www.messenger.com/`. Accessed: 8 March, 2017.

[8] A. Herzberg and H. Leibowitz. Can Johnny finally encrypt? Evaluating E2E-encryption in popular IM applications. In *Sixth International Workshop on Socio-Technical Aspects in Security and Trust (STAST 2016)*, Los Angeles, California, USA, 2016.

[9] M. S. Melara, A. Blankstein, J. Bonneau, E. W. Felten, and M. J. Freedman. CONIKS: Bringing key transparency to end users. In *Twenty-Fourth USENIX Security Symposium (USENIX Security 2015)*, pages 383–398. USENIX Association, 2015.

[10] S. Milgram and E. Van den Haag. Obedience to authority, 1978.

[11] S. Ruoti, J. Andersen, T. Hendershot, D. Zappala, and K. Seamons. Private Webmail 2.0: Simple and easy-to-use secure email. In *Twenty-Ninth ACM User Interface Software and Technology Symposium (UIST 2016)*, Tokyo, Japan, 2016. ACM.

[12] S. Ruoti, J. Andersen, D. Zappala, and K. Seamons. Why Johnny still, still can't encrypt: Evaluating the usability of a modern PGP client, 2015. arXiv preprint arXiv:1510.08555.

[13] J. Sauro. *A practical guide to the system usability scale: Background, benchmarks & best practices*. Measuring Usability LLC, 2011.

[14] J. Sauro and J. R. Lewis. Average task times in usability tests: what to report? In *Twenty-Eighth ACM Conference on Human Factors in Computing Systems (CHI 2010)*, pages 2347–2350. ACM, 2010.

[15] S. Schröder, M. Huber, D. Wind, and C. Rottermanner. When SIGNAL hits the fan: On the usability and security of state-of-the-art secure mobile messaging. In *First European Workshop on Usable Security (EuroUSEC 2016)*, 2016.

[16] S. Sheng, L. Broderick, C. Koranda, and J. Hyland. Why Johnny still can't encrypt: Evaluating the usability of email encryption software. In *Poster Session at the Symposium On Usable Privacy and Security*, Pitsburg, PA, 2006.

[17] A. Sotirakopoulos, K. Hawkey, and K. Beznosov. "I did it because i trusted you": Challenges with the study environment biasing participant behaviours. In *SOUPS Usable Security Experiment Reports (USER) Workshop*, 2010.

[18] J. Tan, L. Bauer, J. Bonneau, L. F. Cranor, J. Thomas, and B. Ur. Can unicorns help users compare crypto key fingerprints? In *Thirty-Fifth ACM Conference on Human Factors and Computing Systems (CHI 2017)*, pages 3787–3798. ACM, 2017.

[19] N. Unger, S. Dechand, J. Bonneau, S. Fahl, H. Perl, I. Goldberg, and M. Smith. SoK: secure messaging. In *Thirty-Sixth IEEE Symposium on Security and Privacy (SP 2015)*, pages 232–249. IEEE, 2015.

[20] E. Vaziripour, M. O'Neill, J. Wu, S. Heidbrink, K. Seamons, and D. Zappala. Social authentication for end-to-end encryption. In *Who Are You?! Adventures in Authentication (WAY 2016)*. USENIX Association, 2016.

[21] Viber. Viber.com. `https://www.viber.com/en/`. Accessed: 8 March, 2017.

[22] WhatsApp. Whatsapp.com. `https://www.whatsapp.com/`. Accessed: 8 March, 2017.

[23] A. Whitten and J. Tygar. Why Johnny can't encrypt: A usability evaluation of PGP 5.0. In *Eighth USENIX Security Symposium (USENIX Security 1999)*, pages 14–28, Washington, D.C., 1999. USENIX Association.

APPENDIX

A. STATISTICAL TESTS

This section contains the details of the statistical tests we ran.

A.1 Success and Failure Rates

This data measures whether the participants were successful in using the authentication ceremony for each application in the second phase of the study. We want to test whether there are any differences between the applications.

Because the data is dichotomous we used Cochran's Q Test and found that the success rate was statistically different for the applications ($\chi^2(2) = 15.429$, p<.0005).

We then ran McNemar's test to find the significant differences among the pairs of applications. As shown in Table 7, and after applying a manual Bonferroni correction for the three tests (requiring p<0.0167), there is a significant difference between WhatsApp and Viber (p=0.008) as well as between Facebook Messenger and Viber (p<0.0005).

		Fail	Success	N	Exact Sig.
		Viber			
WhatsApp	Fail	2	8	48	0.008
	Success	0	38		
		Viber			
Messenger	Fail	0	12	42	0.000
	Success	0	30		
		Messenger			
WhatsApp	Fail	4	2	42	0.109
	Success	8	28		

Table 7: McNemar's test for success and failure

A.2 Task Completion Times

This data measures the time taken by participants to (a) find the authentication ceremony and (b) complete the authentication ceremony, which was only measured in the second phase of the study. We want to know if there is a significant difference in the time to complete these tasks among the three different applications tested—WhatsApp, Viber, and Facebook Messenger.

We first tested for normality using the Shapiro-Wilk test. As Table 8 shows, the data is not normally distributed for any application (p<0.05).

Task	Application	Statistic	df	Sig.
Finding Ceremony	WhatsApp	0.856	38	0.000
	Viber	0.835	46	0.000
	Messenger	0.841	30	0.000
Completing Ceremony	WhatsApp	0.902	38	0.003
	Viber	0.878	46	0.000
	Messenger	0.886	30	0.004

Table 8: Shapiro-Wilk test for task completion times

We next ran the Kruskal-Wallis test, which is a nonparametric test that can determine if there are statistically significant differences between two or more groups. This test retains the null hypothesis that the distribution of task times is the same across the applications for finding the ceremony (p=0.278), but rejects the null hypothesis for completing the ceremony (p=0.043). We next ran pairwise post-hoc tests to determine where the differences occur for completing the ceremony.

As Table 9 shows, we found a significant difference between Viber and WhatsApp for completing the ceremony (p=0.045), with Viber being faster (mean time WhatsApp=8.5 minutes, Viber 6.7 minutes). Note, the significance has been adjusted by the Bonferonni correction for multiple tests.

Task	Comparison	Test Statistic	Std. Error	Std. Test Statistic	Adj. Sig.
Completing Ceremony	Messenger Viber	-12.000	7.702	-1.558	0.358
	Viber - WhatsApp	17.526	7.195	2.436	**0.045**
	Messenger - WhatsApp	5.526	8.016	0.689	1.000

Table 9: Pairwise comparisons from Kruskal-Wallis post-hoc tests for task completion times

A.3 Favorite Rankings

This data measures the system participants selected as their favorite, which was only collected in the second phase of the study. We want to test whether there are any differences between the favorite rankings for each application between the two phases.

We ran a Chi-Square test using the scores for the favorite application. As shown in Table 10, there are no statistically significant differences.

Phase	Favorite WhatsApp	Favorite Viber	Favorite Messenger	Pearson Chi-Square	df	Asym. Sig.
1	9	2	11	2.069	2	0.355
2	15	11	21			

Table 10: Chi-Square test for favorite application ranking

A.4 Trust Scores

We ran a mixed model ANOVA Test because we are interested in seeing the interaction between two independent variables (application and phase). This data is not well suited to a Kruskal-Wallis test because the use of the Likert scale provides too many ties when measuring trust. Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction ($\chi^2(2) = 3.385$, p=.184).

We next examined the results for tests of within-subject effects and found that there is a significant interaction between the application and the study phase (F(2,140)=5.023, p=0.008, partial $\eta^2 = 0.067$).

To determine whether there was a simple main effect for the application, we ran a repeated measures ANOVA on each phase. As shown in Table 11, there was a statistically significant effect of the application on trust for phase 1 (F(2,46)=4.173, p=0.022, partial $\eta^2 = .154$). Note that due to a violation of the sphericity assumption in phase 2, we use the Greenhouse-Geisser correction.

Phase	Mean WhatsApp	Mean Viber	Mean Messenger	df	F	Sig.	η^2
1	4.13	3.58	3.79	2,46	4.173	**0.022**	0.154
2	4.10	4.40	4.17	1.69,79.42	1.843	0.171	0.038

Table 11: Repeated measures ANOVA on each phase

By examining the pairwise comparisons, shown in Table 12, we found that the trust score was significantly lower for Viber as compared to WhatsApp in the first phase (M=0.542, SE=0.180, p=0.19). Note, we use the Bonferroni correction for multiple tests.

Comparison	Mean Difference	Std. Error	Adj. Sig	Lower Bound	Upper Bound
WhatsApp-Viber	0.542	0.180	**0.019**	0.076	1.007
WhatsApp-Messenger	0.333	0.155	0.128	-0.068	0.735
Messenger-Viber	0.208	0.225	1.00	-0.373	0.789

Table 12: Pairwise comparisons from one-way ANOVA on each application, phase 1

To determine whether there was a simple main effect for the study phase, we ran a one-way ANOVA on each application to compare the trust between the two phases. As shown in Table 13, there was a statistically significant difference in trust ratings between the two phases for Viber (F(1,70)=14.994, p<0.0005, partial $\eta^2 = .176$). The mean trust for Viber in the first phase was 3.58, and in the second phase it increased to 4.40.

Application	Mean Phase 1	Phase 2	df	F	Sig.	η^2
WhatsApp	4.13	4.12	1,70	0.007	0.935	0.00
Viber	3.58	4.40	1,70	14.994	**0.00**	0.176
Messenger	3.79	4.17	1,70	2.230	0.140	0.031

Table 13: One-way ANOVA on each application

B. STUDY MATERIALS

This section contains the study materials we used. The interview guide and interview form were used by the study coordinators to ensure that each pair of participants experienced an identical study. The questionnaire was followed by study participants to guide them through the study.

B.1 Interview Guide

Make sure to complete the following steps:

1. When the participants arrive, read them the following:

   Welcome to our secure messaging application study. We are the study coordinators and are here to assist you as needed.

   Before we start the study, we need you to install the following applications: WhatsApp, Facebook Messenger, Viber.

   In this study, the two of you will be in different rooms and will use the applications to communicate with each other. You will each be asked to play the role of another person. I will provide you with information about this person. During the study, please use the provided information and not your own personal information.

   Notice that even you are in separate rooms, you are welcome to ask for meeting, calling or emailing your study partner during the study if you need to complete the study.

   You will be asked to do the task while you are thinking loud and express your feelings or thoughts about each single task that you are doing. During the course of this study we will be recording what is happening in the study room including your any verbal communication with the study coordinators. These recordings will not be seen by anyone beside the researchers and will be destroyed once our research is complete. We will not collect any personally identifying information. Any data, besides the previously mentioned recordings and answers to the study survey, will be deleted automatically upon your completion of the study.

   You will each receive $10 as compensation for your participation in this study. The expected time commitment is approximately 60 minutes. If you have any questions or concerns, feel free to ask us. You can end participation in this survey at any time and we will delete all data collected at your request. A study coordinator will be with you at all times to observe the study and also to answer any questions you may have.

2. Before going to the study rooms, make sure they sign the audio recording consent form.

3. Make sure their phone has enough space for installing the three apps (you can ask them to install the apps before the study starts)

4. Choose one of the available codes for later usage in the study from the following link (a spreadsheet for time slots)

5. Flip a coin and choose one participant to be Person A and one person to be Person B.

6. Take the user with whom you decided to work to the study room. Complete the following setup steps:

   (a) Ask the participant to sit down.

   (b) Start the audio recording using the phones in the lab.

   (c) Read the following instructions to your participant:
   We are going to ask you to do a series of tasks. Once you are done with each step, let the study coordinator know you have finished the task. You will then fill out a questionnaire and go to the next step.
   We need you to think out loud while you are doing the tasks, meaning you are supposed to talk about how you are accomplishing the task and express any feelings you have.
   If you have any questions about the study ask the study coordinator. Remember you are allowed to talk to or meet your friend during the study.
   Please do not forget think loud.

7. On the chromebook, load the survey from Qualtrics

8. Give the code you already selected to the user.

9. Before using each system, the survey will instruct the participant to tell you they are ready to begin the next task.

10. During the course of the task pay attention to what user is doing and fill out one of the attached sheets.

   (a) The user is supposed to think aloud while doing the tasks. If she forgets, gently remind her.

   (b) If the user complains that he is confused, suggest he can consult with his study partner and do not help him to accomplish the task. Try not to instruct the user when they ask questions. Answer them while giving as little information as you can away about the study, but try to remind him that he has a partner who can help him.

   (c) If it takes the pair too long to use one application (10 minutes), then record that as a failure and guide the user to the next task. If you end the task, inform the other study coordinator that you have done so, so that he catches up with you.

11. When the survey is finished, ask the participant about their experience.

   (a) Use the situations you noted while they took the study or interesting things they said on the survey.

   (b) If they had any problems during the study, ask them to use their own words to describe the problem. Ask them how they would like to see it resolved.

12. When the participant is finished, go to meet the other group in your room. Next, ask them the following questions: (If it is applicable)

   (a) You saw QR codes, strings of digits, and maybe NFC communication (touching your phones) as methods for verifying keys. Which one did you prefer and why?

(b) If you were in a different city or state from your friend, how would you verify your friend's key? Would this be convenient?

(c) Some of these applications, like Facebook Messenger let you chat both securely and insecurely. The rest of the applications only let you have secure conversations. Which approach do you prefer and why?

13. Thank the participants for their time. Help them fill out the compensation forms. Send them to the CS office to be compensated.

14. Stop the audio recording. Save the record by time.

15. Fill in your name:

16. Return this form.

B.2 Interview Form

Study Coordinator's Name:
Study Number:
System:
WhatsApp, Viber, FaceBook Messenger
Start Time:
End Time:
Key Verification:

☐ QR Code

☐ Manual verification via phone call

☐ Manual verification in person

☐ Manual verification other:

☐ NFC

☐ Verified successfully

☐ Notices conversation encrypted

Mistakes Made:

☐ The user sends the key or anything related to the key via the application itself

☐ The user sends sensitive data (the credit card number) unencrypted or before doing the identity verification

☐ Does not use an encrypted conversation

☐ Other:

Other:

☐ The user calls, texts or meets his study partner Explain:

☐ The application crashes and needs to be restarted. Explain:

☐ The user expresses any strong feelings toward the task (e.g. how boring or hard or easy it is) Explain:

☐ Other Explain:

C. STUDY QUESTIONNAIRE
Secure Messaging Application Study

1. Please enter the Type.
   ○ A
   ○ B

2. Please enter the code that study coordinator provides for you, here.

3. What is your gender?
   ○ Male
   ○ Female
   ○ I prefer not to answer

4. What is your age?
   ○ 17 and under
   ○ 18-24
   ○ 25-34
   ○ 35-45
   ○ 46-64
   ○ 65 and over
   ○ I prefer not to answer

5. What is the highest degree or level of school you have completed?
   ○ None
   ○ Primary/grade school (2)
   ○ Some high school, no diploma
   ○ High school graduate: diploma or equivalent (e.g., GED)
   ○ Some college, no diploma
   ○ Associate's or technical degree
   ○ Bachelor's degree
   ○ Graduate/professional degree
   ○ I prefer not to answer

6. What is your occupation or major?

7. Mark any of the following options which apply to you.
   ☐ Others often ask me for help with the computer.
   ☐ I often ask others for help with the computer.
   ☐ I have never designed a website.
   ☐ I have never installed software.
   ☐ I have never used SSH.
   ☐ Computer security is one of my job responsibilities.
   ☐ I have taken courses related to computer security, electronic engineering, security, or IT.
   ☐ I often use secure messaging applications such as WhatsApp.
   ☐ I have never sent an encrypted email.
   ☐ I am familiar with cryptography.
   ☐ I understand the difference between secure and non-secure messaging applications.

8. **(Second phase only)** How would you rate your knowledge of computer security?
   ○ Beginner
   ○ Intermediate
   ○ Advanced

9. **(Second phase only)** If you are reading a website, such as CNN, using HTTP, who can see what you are reading?
   ○ Nobody, this is a private connection.
   ○ Your ISP and CNN, but nobody else.
   ○ Any router in between you and CNN.
   ○ Your ISP and nobody else.
   ○ I don't know

10. **(Second phase only)** If you use a regular text messaging application, who can read your text messages?
    - ○ Only the person you send the text message to.
    - ○ The person you send the text message to and the company providing the text messaging service.
    - ○ Anybody who is nearby.
    - ○ Google.
    - ○ I don't know.

11. **(Second phase only)** How can you tell if it is safe to enter your username and password on a website?
    - ○ The website has a good reputation.
    - ○ The website has a privacy statement.
    - ○ There is a lock icon in the URL bar and the URL shows the right host name.
    - ○ The web site is professionally designed.
    - ○ I don't know.

12. **(Second phase only)** What is phishing?
    - ○ Making a fake website that looks legitimate to steal your private information.
    - ○ Hacking someone's computer.
    - ○ Calling someone pretending to be a company to steal their information.
    - ○ Tracking your internet habits to send advertisements.
    - ○ I don't know.

13. **(Second phase only)** What is a public key used for?
    - ○ I do not know what a public key is.
    - ○ To encrypt data for the person who owns the corresponding private key.
    - ○ To setup 2-factor authentication so your password can't be stolen.
    - ○ To identify you to a bank.
    - ○ To protect an application so you know it is safe to use.

14. **(Second phase only)** If you receive a message encrypted with your friend's private key, then you know that
    - ○ Your friend has been hacked.
    - ○ Your friend was the one who sent the message.
    - ○ Everything you send your friend is private.
    - ○ You can't trust what your friend is sending you.
    - ○ I do not know what a private key is.

15. Which of the following applications have you ever used? Select as many options that applies to you.
    - ☐ WhatsApp
    - ☐ ChatSecure
    - ☐ Signal
    - ☐ Telegram
    - ☐ Zendo
    - ☐ SafeSlinger
    - ☐ Allo
    - ☐ FB messenger
    - ☐ iMessage
    - ☐ imo
    - ☐ Skype
    - ☐ Viber
    - ☐ Other

16. What is the main reason why you use these applications (list of applications from previous question) ?

17. Have you ever tried to verify the identity of the person you are communicating with when you are using (list of applications from previous question) ?
    - ○ Yes
    - ○ No
    - ○ Not Sure

18. Have you ever tried to send sensitive information when you use (list of applications from previous question)?
    - ○ Yes
    - ○ No

19. Have you ever had an experience or heard any stories about any secure messaging applications being compromised?
    - ○ Yes
    - ○ No

20. If yes, what story did you hear and what application was it about?

21. **Second Phase Only:**

    Read aloud the following instructions:

    What Is Secure Messaging?
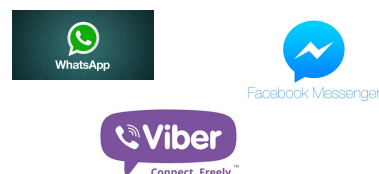
    When you use regular text messaging, your phone company can read your text messages.

    When you use secure messaging apps, you are having a private conversation with your friend.

    Not even the company running the service can see your messages.

But you still need to be careful. A hacker could intercept your traffic.

The bad guy

WhatsApp

To make sure your conversation is secure, these applications assign a "key" to each person.

You need to make sure the key you see is the same key your friend sees.

WhatsApp

Secure messaging apps provide a way for you to compare these keys.

We want to see how well the application helps you do this.

WhatsApp

22. Tell the study coordinator that you are ready for the next task to begin.

Repeat the following block for each of the three applications

23. You would like to send secure text messages to your friend. For example, you might want to ask for a credit card number you left at home, or talk confidentially about a friend who is depressed.

In this study we need you to do the following steps:

For Person A

You are going to be using (WhatsApp/Viber/Facebook Messenger) for secure texting with your friend. This application is designed to help you have a private conversation with your friend.

Your task is to make sure that you are really talking to your friend and that nobody else (such as the service provider) can read your text messages. The application should have ways to help you do this.

We want you to talk and think aloud as you figure this out.

Once you are sure the conversation is secure, ask the other person to send you your credit card number with the following message.

"Hello! Can you send me my credit card number that I left on my desk at home?"

For Person B

You are going to be using (WhatsApp/Viber/Facebook Messenger) for secure texting with your friend. This

application is designed to help you have a private conversation with your friend.

Your task is to make sure that you are really talking to your friend and that nobody else (such as the service provider) can read your text messages. The application should have ways to help you do this.

We want you to talk and think aloud as you figure this out.

Say out loud why you believe you are texting to the right person and why nobody else can read the text messages. Your preference is to figure this out without the other person in the same room, but If you need to visit the other person to do this, you should go ahead and visit them.

Once you are sure the conversation is secure, he/she will ask you to send his/her credit card number through the application. Use the following number in the study: "132542853779"=

24. You will now be asked several questions concerning your experience with (WhatsApp/Viber/Facebook Messenger).

25. **(Second phase only)** Please answer the following questions about (WhatsApp/Viber/Facebook Messenger). Try to give your immediate reaction to each statement without pausing to think for a long time. Mark the middle column if you don't have a response to a particular statement.

- I think that I would like to use this system frequently.
  - Strongly agree
  - Somewhat agree
  - Neither agree nor disagree
  - Somewhat disagree
  - Strongly disagree
- I found the system unnecessarily complex.
  - Strongly agree
  - Somewhat agree
  - Neither agree nor disagree
  - Somewhat disagree
  - Strongly disagree
- I thought the system was easy to use.
  - Strongly agree
  - Somewhat agree
  - Neither agree nor disagree
  - Somewhat disagree
  - Strongly disagree
- I think that I would need the support of a technical person to be able to use this system.
  - Strongly agree
  - Somewhat agree
  - Neither agree nor disagree
  - Somewhat disagree
  - Strongly disagree
- I found the various functions in this system were well integrated.
  - Strongly agree
  - Somewhat agree

- ○ Neither agree nor disagree
- ○ Somewhat disagree
- ○ Strongly disagree

- I thought there was too much inconsistency in this system.
  - ○ Strongly agree
  - ○ Somewhat agree
  - ○ Neither agree nor disagree
  - ○ Somewhat disagree
  - ○ Strongly disagree
- I would imagine that most people would learn to use this system very quickly.
  - ○ Strongly agree
  - ○ Somewhat agree
  - ○ Neither agree nor disagree
  - ○ Somewhat disagree
  - ○ Strongly disagree
- I found the system very cumbersome to use.
  - ○ Strongly agree
  - ○ Somewhat agree
  - ○ Neither agree nor disagree
  - ○ Somewhat disagree
  - ○ Strongly disagree
- I felt very confident using the system.
  - ○ Strongly agree
  - ○ Somewhat agree
  - ○ Neither agree nor disagree
  - ○ Somewhat disagree
  - ○ Strongly disagree
- I needed to learn a lot of things before I could get going with this system.
  - ○ Strongly agree
  - ○ Somewhat agree
  - ○ Neither agree nor disagree
  - ○ Somewhat disagree
  - ○ Strongly disagree

26. I trust this application to be secure.
    - ○ Strongly agree
    - ○ Somewhat agree
    - ○ Neither agree nor disagree
    - ○ Somewhat disagree
    - ○ Strongly disagree

27. Have you managed to verify the identity of your friend correctly?
    - ○ No
    - ○ Yes
    - ○ Not sure

28. Please explain why do you think you have (or have not) verified the identity of your friend.

29. Who do you think can read your message except you and your friend?
    End of the repeated block

30. You have finished all the tasks for this study. Please answer the following questions about your experience.

31. Which system was your favorite?
    - ○ WhatsApp
    - ○ Viber
    - ○ FaceBook Messenger
    - ○ I didn't like any of the systems I used

32. Please explain why.

33. Which of the following applications have you ever used for secure communication? Select as many options that applies to you.
    - ☐ WhatsApp
    - ☐ ChatSecure
    - ☐ Signal
    - ☐ Telegram
    - ☐ Zendo
    - ☐ SafeSlinger
    - ☐ Allo
    - ☐ FB messenger
    - ☐ iMessage
    - ☐ Skype
    - ☐ imo
    - ☐ Viber
    - ☐ Other

34. Please answer the following question. Try to give your immediate reaction to each statement without pausing to think for a long time. Mark the middle column if you don't have a response to a particular statement.

    It is important to me to be able to have private conversations with my friends and family using secure applications (like WhatsApp).

    - ○ Strongly disagree
    - ○ Disagree
    - ○ Neither Agree nor Disagree
    - ○ Agree
    - ○ Strongly Agree

35. Did you know about encryption before attending this study?

36. Are you willing to participate in a follow up study? If so, please leave your name and phone number with the study coordinator.