Poster: Can Johnny Authenticate?

Elham Vaziripour, Ray Clinton, Justin Wu, Mark O'Neill,
Jordan Whitehead, Scott Heidbrink, Kent Seamons, Daniel Zappala
Computer Science Department,
Brigham Young University
elhamvaziripour@byu.edu, bigblue0@studentbody.byu.edu, justinwu@byu.edu, mto@byu.edu,
jaw@byu.edu,sheidbri@byu.edu, seamons@cs.byu.edu, zappala@cs.byu.edu

I. Introduction

Over the past several years, disclosures of widespread surveillance of Internet traffic and security breaches have brought increased attention to using end-to-end encryption for Internet communication. Of particular note is the now widespread use of secure messaging applications such as WhatsApp, Signal, Facebook Messenger, Viber, and so forth. These applications have generally chosen to favor usability over security, and indeed recent research shows that these applications are rarely adopted for security or privacy reasons, and are instead primarily adopted due to peer influence [2].

However, the effective security being provided by secure messaging applications depends heavily on users completing an authentication ceremony in order to establish trust[8], and the evidence to date suggests users are unable to do this [6], [4], [3], [7], particularly when keys change due to a reinstall or an attack. For example, Schroder et al. studied the usability and security of Signal, showing that users were vulnerable to active attacks due to usability problems and incomplete mental models of public key cryptography [6]. Herzberg and Leibowitz examined the usability of WhatsApp, Viber, Telegram, and Signal, finding that most users were unable to properly authenticate, both in an initial authentication ceremony and after a key reset [4]. Dechand et al. study the usability of textual key verification methods, finding that users are more resistant to attacks when using sentence-based encoding as compared to hexadecimal, alphanumeric, or pure numeric representations [3]. Tan et al. examined usability of eight various fingerprint representations under realistic conditions, finding that graphical representations are more susceptible; however they are easy and quick to use [7].

In this work, we conduct a between-subjects study to analyze 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 applications: WhatsApp, Viber, and Facebook Messenger. 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 find differences in key verification success rates when users are provided with instruction concerning only the need for authentication versus when they are also told about the importance of comparing keys. By observing participants as they use the applications, we identify the methods they choose for completing the authentication ceremony and note common mistakes and user grievances.

II. 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.

In the first phase of our study, we asked 12 pairs of users to complete a scenario where one participant sends 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 (confidentiality). Participants were told the application would help them accomplish these goals.

In the second phase of the study, we asked 24 pairs of users to complete the same scenario, with the same instructions. However, before completing the task, the participants viewed an instructional set of slides. 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. The slides did *not* show how to find or complete the authentication ceremony in any of the applications.

Upon beginning the study, participants answered a series of questions on a survey, covering demographics, their past experience with secure messaging applications, 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 app, 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 were completed, participants were then

asked which of the three apps was their favorite and why. In addition to quantitative results we also collected qualitative data, requesting participants to think aloud and explain why they think they completed the task successfully.

Our sample population 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.

III. Results

In the first phase of the study, despite the instruction about potential threats, only 2 of the 12 pairs experienced some success in locating and completing the authentication ceremony. Participants who did not succeed in locating the authentication ceremony used a variety of ad hoc methods for authentication. The most common methods used were calling to speak with the other party, and verifying them visually, by recognizing their voice, and by asking questions that depend on shared knowledge. Note that these ad hoc rules indicate participants had in mind a particular threat model, an impersonation attack or physical access to their phones.

In the second phase of the study, the success rate for completing the authentication ceremony was drastically higher. Overall, the success rate was 78% across all participant pairs and the three applications. Successes indicate that participants identified and compared keys in some fashion. 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.

When using WhatsApp, the most-selected authentication method, by 46% of participants, was scanning the QR code of the key fingerprint in person, which is a method unique to WhatsApp. Otherwise, they chose less secure methods, such as sending the key through the application. 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 these data are provided to the user. As a result, all pairs who successfully completed the Viber ceremony, 96% of participants, 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 app without performing this procedure. Finally, the most-selected secure method for authentication Facebook Messenger (FBM), for 29% of participants, was reading the key data in person. However, more mistakes were made in this application, such as by exchanging the key via the application.

We examined the time taken to find and use the authentication ceremony in each application. A major takeaway from the timing data is that key discovery and key verification both require substantial time for all three applications. On average, across all applications, locating the ceremony required 3.5 minutes and completing the ceremony 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.

During the second phase of our study, participants evaluated each application using the System Usability Scale (SUS). The applications' overall SUS scores fall within the C range, landing somewhere within the 41st to 59th percentile [1].

Finally, we asked users about how much they trusted the applications. Trust generally increased from the first phase to the second phase, once users knew more about the security provided by the applications. However, trust was often based simply on the reputation of a company offering the application, and even after having some of the security concepts explained, users remarked that they had no way to truly gauge the promises made by the applications to secure their communication.

IV. Conclusion

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 have the right 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. Another area for future work is improving the authentication ceremony so that it does not take so long to complete. Finally, better methods are needed for obtaining public keys without relying on a single trusted party. Some possibilities include using CONIKS [5] or social authentication [9].

References

[1] Bangor, A., Kortum, P. T., and Miller, J. T. An empirical evaluation of the System Usability Scale. *International Journal of Human–Computer Interaction 24*, 6 (2008), 574–594.

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

[3] Dechand, S., and Schürmann, D. An empirical study of textual key-fingerprint representations. In *USENIX Security Symposium* (2016), USENIX.

[4] Herzberg, A., and Leibowitz, H. Can Johnny finally encrypt? Evaluating E2E-Encryption in popular IM applications. In *International Workshop on Socio-Technical Aspects in Security and Trust* (2016), USENIX.

[5] Melara, M. S., Blankstein, A., Bonneau, J., Felten, E. W., and Freedman, M. J. CONIKS: Bringing key transparency to end users. In *USENIX Security Symposium* (2015), USENIX.

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

[7] Tan, J., Bauer, L., Bonneau, J., Cranor, L. F., Thomas, J., and Ur, B. Can unicorns help users compare crypto key fingerprints? In *SIGCHI Conference on Human Factors in Computing Systems (CHI)* (2017), ACM.

[8] Unger, N., Dechand, S., Bonneau, J., Fahl, S., Perl, H., Goldberg, I., and Smith, M. SoK: Secure messaging. In *IEEE Symposium on Security and Privacy (SP)* (2015), IEEE.

[9] Vaziripour, E., ONeill, M., Wu, J., Heidbrink, S., Seamons, K., and Zappala, D. Social authentication for end-to-end encryption. In *"Who Are You?!" Adventures in Authentication Workshop* (2016), USENIX.