Errata Slip

Proceedings of the 2017 Symposium on Usable Privacy and Security

In the paper "Is that you, Alice? A Usability Study of the Authentication Ceremony of Secure Messaging Applications" by Elham Vaziripour, Justin Wu, Mark O'Neill, Ray Clinton, Jordan Whitehead, Scott Heidbrink, Kent Seamons, and Daniel Zappala, *Brigham Young University* (Thursday session, "Authentication", pp. 29–47 of the Proceedings), the reported results indicate a statistical difference was found in the time to find the authentication ceremony in the secure messaging applications that were studied. In verifying the statistics reported in the paper, the authors found that this was calculated incorrectly and the difference is not significant. This affects sections 6.3 and Appendix A.2 of the paper.

Original text from section 6.3

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 both finding the ceremony ($p = 0.031$) and 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 Facebook Messenger and WhatsApp for finding the ceremony ($p = 0.030$), with Facebook Messenger being faster (mean time, Facebook Messenger=2.5 minutes, WhatsApp=3.7 minutes). We also 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).

Corrected text

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).

Original text and tables from section A.2

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	0.902	38	0.003
	Viber	0.878	46	0.000
	Messenger	0.886	30	0.004
Completing Ceremony	WhatsApp	0.856	38	0.000
	Viber	0.835	46	0.000
	Messenger	0.762	30	0.000

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 rejects the null hypothesis that the distribution of task times is hte same across the applications, for both finding the ceremony (p=0.031) and completing the ceremony (p=0.043). We next ran pairwise post-hoc tests to determine where the differences occur.

As Table 9 shows, We found a significant difference between WhatsApp and Facebook Messenger for finding the ceremony (p=0.029), with Facebook Messenger being faster (mean time, WhatsApp=3.7 minutes, Facebook Messenger=2.5 minutes). We also found a significant difference between Viber and WhatsApp for completing the ceremony (p=0.021), 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.
Finding Ceremony	Messenger - Viber	14.887	7.616	1.955	0.152
	Viber - WhatsApp	5.492	7.114	0.772	1.000
	Messenger - WhatsApp	20.379	7.926	2.571	**0.030**
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

Corrected text and tables

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