Privacy in the Amazon Alexa Skills Ecosystem

Abdulaziz Alhadlaq
University of Southern California
alhadlaq@usc.edu

Jun Tang
University of Southern California
juntang@usc.edu

Marwan Almaymoni
University of Southern California
almaymon@usc.edu

Aleksandra Korolova
University of Southern California
korolova@usc.edu

1. ABSTRACT

Echo is a new type of device placed in people's homes and powered by Amazon's voice service, Alexa. In 2015 Amazon has introduced Alexa Skills Kit, a set of APIs that enable developers to build voice-driven capabilities, called Skills, for Alexa. In this work, we study the privacy-related choices made by Amazon when establishing the Skills ecosystem and the developers' current privacy practices within it. Specifically, we analyze the privacy policies of all Alexa Skills, and find that 75% do not have one. Furthermore, even among those Skills for whom a privacy policy is required by Amazon's policy, 3.5% do not have a valid one and 70% are not customized to Alexa. We then describe and discuss the design choices made by Amazon in exposing user information to Skills developers. Since the ecosystem is very new and the privacy challenges it poses may be distinct from those in Apple's App Store and Android's Play Store, our goal is to raise awareness of the current design choices, and facilitate discussion that could help set meaningful privacy standards for this new domain of growing importance.

2. INTRODUCTION TO ALEXA SKILLS

Skills are the new voice-driven capabilities developers can create for the cloud-based Alexa service that powers Amazon devices, such as Echo. After approval by Amazon, they are published in the Alexa Skills Store [1] and as of February 2017, more than 10,000 Skills were available [2]. The entry for each Skill in the Alexa Skills Store contains its name, developer, user rating, a description of Skill's capabilities, a guidance rating, and a Skill Details section, which may include links to the (optional) Privacy Policy and Terms of Use.

To use a Skill, a user simply invokes it via a voice command on their Echo device; e.g., "Alexa, let me talk to <invocation name>". Such invocations do not require prior installation or action on the user's part as the Skills themselves are not installed on the Echo; instead, the device acts as an agent between the user and the cloud-based backend of the Skill. Thus, when a Skill is invoked, the user is not explicitly told of a privacy policy or asked to agree to a privacy policy or terms of use. A user who wishes to inspect such policies must access them manually online.

Account Linking: Some Skills require the user to link their Alexa account to an existing account on another service. For example, when using the Uber Skill on Alexa for the first time, the user will be required to link to their existing Uber account, which allows the Skill to identify the user whenever they access it. To perform linking, the user is required to use the Amazon Alexa mobile companion app and manually link the account.

3. SKILLS' PRIVACY POLICY ANALYSIS

In May 2017, we performed a crawl using Selenium of the Alexa Skills website to extract Skills' description and privacy policies, which resulted in 11,827 Skills. Our analysis results can be found in Tables 1, 2, and 3.

Not all Skills have a privacy policy as Amazon does not require one unless the Skill performs *Account Linking*. In fact, we found that ~75% of all Skills do not have a privacy policy. Furthermore, as shown in Table 1, even among the 2,860 Skills with a link to a privacy policy, 269 are unreachable due to 404 and 500 HTTP errors, timeout errors, and access denied errors. Additionally, 35 of the privacy policy links direct to pages that do not appear to be privacy policies at all and instead direct to pages such as the main page of the developer's website. Thus, among all Skills with Privacy Policy links, the fraction of invalid privacy policy pages (i.e., pages that are either inaccessible or not privacy pages) is almost 11% (3.5% for those that require Account Linking). This shows that both Amazon's and Skill developers' attitude towards privacy policies is quite sloppy.

Furthermore, most of the privacy policies linked are the ones used for the developer's website or mobile app, and are not tailored for Alexa. Only 407 (or 14% of valid privacy policies) contain one of "alexa", "echo", or "amazon skill".

Table 2 presents privacy policy statistics broken down by the rating of a Skill. Surprisingly, the 5-star rated Skills have the lowest rate of inclusion of a privacy policy (18%), whereas the 1-star rated ones have the highest (26%).

Table 1. A Breakdown of Privacy Policy Presence Depending on Skills' Account Linking Behavior

	No account linking	Account linking	Total
No privacy policy	8,967	0	8,967
Error page [*]	252	17	269
Wrong page	28	7	35
References Alexa [**]	227	180	407
No reference of Alexa	1,675	474	2,149
Total Skills	11,149	678	11,827

[*] Pages with timeout, access denied, 404 or 500 HTTP errors.
[**] Pages that contain keywords: "alexa", "echo", "amazon skill".

Table 2. Privacy Policy Statistics by Skill's Star Rating

Rating	Has a Privacy Policy	Account Linking	Wrong page	Error page	Total
1 Star	1902	217	11	249	7195
2 Star	91	26	1	1	491
3 Star	266	138	4	4	1160
4 Star	282	142	7	6	1186
5 Star	319	154	12	9	1795

Table 3. Privacy Policy Statistics by Skill Category

Category	Has a Privacy Policy	Account Linking	Wrong page	Error page	Total Skills in Category
Business	103	44	2	4	288
Communication	21	16	0	0	46
Connected Car	17	17	0	1	23
Education	109	21	4	2	1852
Food	51	34	0	2	299
Game	92	13	2	4	2749
Health	106	45	0	3	363
Lifestyle	210	92	2	6	1038
Local	40	16	0	2	254
Movie	17	5	0	0	178
Music	63	13	11	2	358
News	1651	10	6	238	2689
Novelty	25	4	2	1	947
Productivity	107	67	0	1	447
Shopping	15	9	0	0	52
Smart home	362	347	6	5	406
Social	39	28	1	0	115
Sports	28	5	0	1	332
Travel	44	11	0	4	264
Utilities	30	15	0	2	187
Weather	26	4	0	0	542

Table 3 presents statistics by Skill category. The categories that stand out with a relatively high rate of privacy policy inclusions are "Connected Car", "Smart Home" and "News". For the first two, it is due to most of them also requiring Account Linking. For the "News" category, it may be due to the established news sites linking to their established privacy policies. The categories that stand out with a relatively low rate of privacy policy inclusions are "Education", "Game", "Novelty", and "Weather". The relatively high error page number in "News" and "Music" categories are due a large number of Skills by the same developer.

4. OTHER ALEXA PRIVACY CHOICES

Audio availability: Although users interact with Skills using voice commands, the audio is not made accessible to the developers. Instead, all voice commands are transcribed and sent to the developers as text. This design choice by Amazon has several privacy benefits: first, Skill developers cannot create a unique user fingerprint using the rich audio data. Second, the voice data cannot be used to infer user's age, gender, emotional state, etc., unless this information is explicitly conveyed in the choice of vocabulary (e.g., using gendered verbs and expletives).

userId: The userId is a unique identifier given by Amazon to each user. Amazon assigns a different userId for each Skill used, so, in a win for privacy, the userId's usefulness for cross-Skill tracking is limited. However, the process for re-setting the per-Skill userId is cumbersome: one needs to disable and re-enable it.

Location and List Access: Developers can configure their Skill to request access to the address of the user's device at the "Full Address" or "Country & Postal Code Only" levels. They can also configure it to read and write to user's Alexa lists, such as shopping or to-do. A user enabling such a Skill is prompted for consent with a permissions card in the Alexa app. List permissions apply to all of the user's lists.

For both Location and List Access, Amazon is following the permission model initially adopted for apps in Android and iOS. Although permissions can be revoked in the Skills section of the Alexa app, the process for doing so is per-Skill and is very cumbersome. Moreover, unlike with Account Linking, Amazon is not requiring Skill developers requesting permissions to include a privacy policy. We believe that unless Amazon incorporates innovations proposed by the academic community, its Skills ecosystem will soon suffer from the same woes as other app ecosystems, such as user inability to meaningfully manage Skill permissions without investing an exorbitant amount of time and mismatch between stated and actual data use by developers.

5. RELATED WORK

There's an extensive body of work for privacy analysis of mobile applications (e.g., [3]), and more recently, for security of IoT [4]. To our knowledge, this is the first study of Alexa Skills privacy.

6. REFERENCES

[1] Amazon Alexa Skill Store: https://amazon.com/skills

[2] Alexa Now Has Over 10,000 Skills Available, Amazon's Developer Blog, Feb 23, 2017.

[3] S. Zimmeck, Z. Wang, L. Zou, R. Iyengar, B. Liu, F. Schaub, S. Wilson, N. Sadeh, S.M. Bellovin, J.R. Reidenberg, "Automated Analysis of Privacy Requirements for Mobile Apps" in Proceedings of Network and Distributed System Security Symposium (NDSS), 2017.

[4] E. Fernandes, J. Jung, A. Prakash. "Security analysis of emerging smart home applications." in Proceedings of the IEEE Security and Privacy Symposium (S&P), 2016.