Hey Alexa, is this Skill Safe?
Taking a Closer Look at the Alexa Skill Ecosystem

Abstract
Amazon's voice-based assistant, Alexa, enables users to directly interact with various web services through natural language dialogues. It provides developers with the option to create third-party applications (known as Skills) to run on top of Alexa. While such applications ease users' interaction with smart devices and bolster a number of additional services, they also raise security and privacy concerns due to the personal setting they operate in. This paper aims to perform a systematic analysis of the Alexa skill ecosystem. We perform the first large-scale analysis of Alexa skills, obtained from seven different skill stores totaling to 90,194 unique skills. Our analysis reveals several limitations that exist in the current skill vetting process. We show that not only can a malicious user publish a skill under any arbitrary developer/company name, but she can also make backend code changes after approval to coax users into revealing unwanted information. We, next, formalize the different skill-squatting techniques and evaluate the efficacy of such techniques. We find that while certain approaches are more favorable than others, there is no substantial abuse of skill squatting in the real world. Lastly, we study the prevalence of privacy policies across different categories of skill, and more importantly the policy content of skills that use the Alexa permission model to access sensitive user data. We find that around 23.3% of such skills do not fully disclose the data types associated with the permissions requested. We conclude by providing some suggestions for strengthening the overall ecosystem, and thereby enhance transparency for end-users.

Citation

Published Paper