Tufts University

Department of Computer Computer Science

COMP116: Introduction To Computer Security
Final Project

Investigation Of Amazon Alexa's Explicit Invocation Policy

Author:
Holt Spalding

Supervisor:
Ming Chow

December 12, 2018

Abstract

In 2014, Amazon released the first version of its voice-powered personal assistant, the Amazon Echo Dot. Since then, the popularity of Alexa devices and similar devices has grown rapidly, with nearly 16% of US adults owning an Amazon or Google brand personal assistant. While the popularity of the Alexa has grown significantly over the last few years, so too has consumer paranoia surrounding the security of the product. And consumers have good reason to be paranoid. Just last year, researchers at the University of Indiana found two major security vulnerabilities in the Amazon Alexa which have given rise to a new form of cyber attack known as voice squatting. Last year Amazon released their so-called skill store, a marketplace allowing third-party developers to publish their own native Alexa applications. In a voice squatting attack, opportunistic hackers will publish malware that bears a name homophonous to one of the more popular applications. For example, a hacker could dub their malware Capital Won, which may lead to users to unintentionally reveal sensitive financial information to a malicious application which they believe to be the Capital One app. This paper seeks to assess the extent to which voice squatting can still be exploited on the skill store. Furthermore, this paper seeks to uncover any other vulnerabilities in the Alexa which may directly result from the linguistic ambiguity in speech.

1 A Brief Note

This paper represents just a small part of a larger project to uncover security vulnerabilities present in Alexa-enabled IoT devices that result directly from their use of human voice commands to control software. The project was originally designed with the expectation that not all aspects of it could be completed in only a few weeks time. Amazon is one of the worlds largest companies and they make great strides to hide the proprietary systems that power their products from the eyes of consumers. To expect there wouldn't be any roadblocks along the way would have been unwise. While I think this represents just a small contribution to currently minuscule body of research surrounding this topic, I do not see it as insignificant or unsuccessful. At the very least I view this paper as a call to action for people in the cybersecurity world to invest more time and research into the issues mentioned below. It incredibly important that we hold companies like Amazon to account for the security risks they expose their user base to on a daily basis. In developing this project I also had the opportunity to learn a lot about how Voice-activated Personal Assistants (VPAs) actually function, and I hope through this paper I can provide at least a sliver of that knowledge to anyone interested in investigating the Alexa further. Expect to see a follow up in the near future. Thank you.

2 Background

Before diving into Alexa's explicit invocation policy, it's important you first have some background on Alexa products and how they function. Then, I will provide a brief explanation of the findings of my experiment on explicit invocation.

2.1 Virtual Personal Assistant Systems

Amazon and Google have become major players in the smart speaker market over the last 4 years. Amazon released its first Virtual Personal Assistant (VPA) smart speaker, the Amazon Echo, in late 2014 and since then they've grown to control over 60% of the smart speaker market share in the US.[1] According to the latest quarterly research from Strategy Analytics, Amazon dominates the smart speaker market globally as well, controlling nearly

40% of the global market share (though this has been changing over the last year since the release of Google's propriety "Google Home" device)[2]. Based on this data, experts and industry insiders have speculated that almost 50 million units of the Amazon echo have been sold in the US alone[3,4]. Additionally, Amazon has integrated Alexa into countless "smart" IoT products from other vendors like the Sonos-brand smart speaker, various power outlets, security cameras, thermostats, doorbells, and even a smoke detector[5]. The Alexa is one of the most commercially successful IoT recording devices ever put out on the market, and as such, its important that we as a society remain skeptical of Amazon's claims of its security.

2.2 Alexa Skills

The capabilities of Amazon's Alexa products are extensible thanks to Amazon's digital online market place known as *The Skill Store*. The Skill Store is a platform where third party vendors can develop and publish software using Amazon's *Alexa Skill Kit*[6] that can then be run on any Alexa-enabled device. Today there are over 30,000 skills on the store representing a few thousand brands across different industries. Popular brans featured on the Skill Store Amex, Capital One, Fox Cable Network, NBC, Sirius XM, Geico, and Kayak just to name a few[7]. Anyone can develop and publish Alexa software through the Skill Store for free with only an Amazon account. Before skills are published, they must undergo a review process at Amazon that usually lasts only a day or two. Amazon has not revealed anything about its internal review process, one of the major roadblocks in investigating this topic.

2.3 Alexa's Interaction Model & Skill Invocation

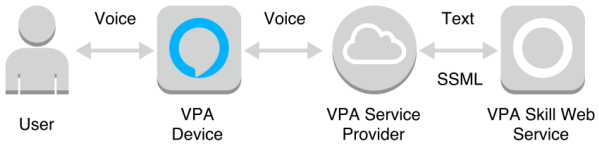

Figure 1: From source 8

The above figure illustrates the *Alexa Interaction Model* [9] which defines the protocol by which users communicate with third-party skills through an Alexa-enabled device. When an Alexa user attempts to interface with a new or existing application on their device, they do this by means of *skill invocation*. New skills can be invoked either implicitly or explicitly. In explicit invocation, users can automatically enable a new skill on their devices simply by saying "Alexa, open" or "Alex, start" and then referring to the new skill by its *wake-word* (also known as an *invocation name*. The wake-word for the "Jeopardy!" skill, a popular game on the Skill Store, is of course "jeopardy" (note that this wake-word may bear no resemblance to the name of the skill itself). Alternatively, skills can be invoked implicitly by a user's Alexa device whenever it believes the conversation demands it. For example, when asking the Alexa for tomorrow's weather forcast, the Alexa will automatically invoke the Amazon-created *Weather* skill. Based on my research and personal experience with the Alexa, as of right now it appears as though the "custom skills" published to the Skill Store can only be enabled by means of explicit invocation, although this requires further research.

When explicitly invoking a custom skill, the user's request is first sent to an AWS server where a speech recognition system converts it to text, determines the skill the user wishes to invoke, and then sends the text along with some other metadata to the third-party application's private server[9]. User requests are handled server-side by a so-called *intent handler*, meant to map user voice requests to *intents* which are essentially just functions which determine the appropriate response or action to any given voice command. Default intents include the *WelcomeIntent*, *HelpIntent*, etc. When developing a skill, you must provide Amazon a set of intents you want your skill to handle as well as a set of sample utterances you would want a users to say in order to activate each intent. Amazon can then presumably extrapolate from those sample utterances you've provided in order to provide users a wide range of ways to say any given command in your skill.

Once the proper response has been formulated, the third-party server sends a response back to the AWS server either in plaintext or something known as Speech Synthesis Markup Language (which also allows embedded mp3s)[10] so a TTS system can send back an audible response to the user.

3 The Vulnerabilities

3.1 Voice Squatting

The Alexa Interaction model has been severely under-studied and very little is known about its potential vulnerabilities. One of the few major studies on the Alexa Interaction model was published earlier this year by researchers at the University of Indiana Bloomington[8]. In that study, the researchers developed a new form of attack on VPA devices known as a "voice-squatting" or "voice masquerading" attack. Voice squatting is a form of phishing attack which exploits the lexical ambiguities in English in order to trick users into unintentionally invoking malicious skills published by a remote adversary. For example, a user may ask their Alexa device to open the "Capital One" banking app, causing their device to instead open the malicious "Capital Won" app. Alternatively, an attacker could also create a malicious skill whose wake-word mirrored some everyday phrase like "Weather please", or in requesting the Capital One app, "Capital One please." The researchers were able to publish 4 new skills to the Skill Store which exploited this vulnerability, all of which garnered a fair amount of traffic.

The risk potential surrounding voice squatting attacks is pretty high... at least far higher than the Amazon Skill Store would seem to reflect. Not only is there the potential for users to reveal sensitive information to malicious actors, but it also provides the perfect framework for eavesdroppers to take out their attacks. In a study published by Checkmarx earlier this year, researchers were able to design a fairly simple skill which posed as a calculator app that was actually capable of recording users even after they had moved on to another application[11].

4 My Experiment

4.1 The Purpose

Amazon has claimed to be working closely with the Indiana University researchers that discovered this voice squatting vulnerability in order to mitigate the problem, although it remains to be seen what steps Amazon has actually taken, if any at all. The overall goal of this larger project is to determine if I can recreate the results of the University of Indiana researchers in order to asses the extent to which this problem has been resolved. Un-

fortunately, my attempts to develop Amazon skills which recreate this voice squatting attack are still undergoing the review process and are yet to be published. Amazon has not officially cited any reason why they've yet to publish my skills, but according to an email correspondence I've had with an Amazon employee I believe it is due to unrelated reasons regarding the status of my AWS account. I hope to expand on this report in the coming year and I will continue to try pushing my skills through the review process. Despite this small setback, I was still able to produce some very interesting findings surrounding explicit invocation. In this experiment, I surveyed the Skill Store to find skills that lacked unique invocation names and then performed a small informal user-study to determine the policy by which Alexa chooses between two similarly-named skills.

4.2 Methodology

I gathered 20 students from Tufts University to act as participants in this experiment, 10 male and 10 female. 16 of the 20 participants were native English speakers. In addition, I found 613 skills that bore just one of 50 invocation names. Some invocation names, like *cat facts*, were shared by almost 100 skills. The mean number of skills sharing the same invocation name was around 23 and the median was 8. In the study, I asked each participant to "explicitly invoke" each invocation name two different times... first by saying "Alexa, open..." and then the invocation name, and then by saying "Alexa, start..." and then the invocation name. I then used my Alexa Skill Kit to determine which skill was actually invoked each time. Also note that I used an Amazon Echo Dot Generation 3 for this experiment, although the model should have no bearing on the results.

4.3 Findings

Through this short informal study I actually found a few very interesting things that will absolutely serve the larger project going into the future. The first thing I found is that there appears to be no apparent or explicit restriction on a skill's chosen invocation name. There are thousands of apps bearing the exact same invocation name on the Skill Store. Some skills even bear invocation names very similar to popular brands and skills like "jeopardy please", however I was not able to find any skills with the *exact* same invocation name as some of the more popular brands on the Skill Store.

The next thing I observed is that Alexa devices appear to employ a random policy when choosing between skills that bear the same invocation name. I was expecting popularity to play a role in determining which skill is invoked on any given ambiguous invocation, but it appears that's not the case at all. All participants were collectively able to invoke about half of the unique "cat fact" skills. Even having a skill already enabled on your device with the same invocation name did not seem to effect the invocation policy at all.

Finally, and I think most importantly, I found that only 11% of the skills asked explictly for user permission before enabling themselves on the Alexa device. Most of the time, skills would just launch as soon as the invocation name is heard. This included some of the more major brands and banking apps like Amex.

5 Conclusion

Collecting this data was less trivial than I had expected since there is currently no formal means by which it can be collected. As of right now, I think there are three primary steps that need to be taken in order to make the Alexa safer: 1) Amazon should generally be more transparent about the inner workings of the Alexa 2) the Alexa should require each skill bear a unique invocation name and 3) security researchers should spend more time investigating voice squatting attacks. My research and the findings of my user study will come in handy as the project develops over the next year or so.

6 References

1. Voice Shopping US Consumer Adoption and Attitudes 2018 Report. (2018, June 01). Retrieved from https://voicebot.ai/voice-shopping-us-consumer-adoption-and-attitudes-2018-report-pre-register/

2. Strategy Analytics: Google Closes Gap on Amazon in Global Smart Speaker Market in Q2 2018. (2018, August 13). Retrieved from https://news.strategyanalytics.com/press-release/devices/strategy-analytics-google-closes-gap-amazon-global-smart-speaker-market-q2-201

3. Smart speaker installed base to hit 100 million by end of 2018. (n.d.). Retrieved from https://www.canalys.com/newsroom/smart-speaker-installed-base-to-hit-100-million-by-end-of-2018 Canalys

4. The Smart Audio Report, Spring 2018. (n.d.). NPR Edition Research Retrieved from https://www.nationalpublicmedia.com/smart-audio-report/latest-report/

5. Onelink. (n.d.). Retrieved from https://onelink.firstalert.com/

6. KUMAR , A., GUPTA , A., CHAN , J., TUCKER , S., HOFFMEISTER , B., AND DREYER , M. Just ask: Building an architecture for extensible self-service spoken language understanding. arXiv preprint arXiv:1711.00549 (2017).

7. Kinsella, B. (2018, March 22). Amazon Alexa Skill Count Surpasses 30,000 in the U.S. Retrieved from https://voicebot.ai/2018/03/22/amazon-alexa-skill-count-surpasses-30000-u-s/ voicebot.ai

8. Zhang, N., Mi, X., Feng, X., Wang, X., Tian, Y., Qian, F. (2018). Understanding and Mitigating the Security Risks of Voice-Controlled Third-Party Skills on Amazon Alexa and Google Home. arXiv preprint arXiv:1805.01525.

9. https://developer.amazon.com/docs/alexa-voice-service/interaction-model.html

10. Ssml. https://www.w3.org/TR/speech-synthesis11/.

11. Rubens, A. (2018, April 29). Eavesdropping with Amazon Alexa. Retrieved from https://www.checkmarx.com/2018/04/25/eavesdropping-with-amazon-alexa/