Measuring How Much IoT Devices Upload via Traffic Analysis

Enze Liu, TJ Smith, Zesen Zhang
University of California, San Diego
{e7liu,tjs003,zez003}@eng.ucsd.edu

ABSTRACT

We examine the problem of how much voice data is being streamed to the server by Echo. Echo is Amazon's smart speaker that has an always-on microphone. Conceptually, it should start recording after hearing a certain wake word (Alexa, Echo, etc.) and then send the recorded audio to its server for further processing. Previous work has largely focused on stopping the adversaries from inferring encrypted user data being transmitted. To our best knowledge, little has been done to verify the fundamental assumption that Echo should only start to record and trasmit data after hearing the wake word. **[[what else is not done by the others]]**?

In this work, we performed measurements that shed lights on these issues. More specifically, we provided answers to the following questions: 1. Is any conversation before the wake word being streamed to the server? 2. **[[what else did we measure in this work?]]**. To achieve this, we pre-recorded a piece of audio that contained an Echo command ("Alexa, where is New York City?") as well as irrelevant conversations prefixing and suffixing the Echo command. Different segments of this audio were played over and over again in a controlled environment, and all packets transmitted by Echo were recorded. After data cleaning and carefully analyzing the traffic data, we are able to reveal **[[X]]** key findings. First, we confirmed that normally Echo would not stream any conversation that happened before the wake word. That said, we did observe that, sporadically, Echo would transmit a non-trivial amount of data during an ongoing conversation that did not contain the wake word. **[[What else did we find]]**.

1 INTRODUCTION

Physical devices connected to the internet, also known as Internet-of-things (IoT), have gained popularity in recent years. Among numerous types of IoT devices, smart speakers with voice assistants, such as Amazon Echo and Google Home, are widely seen in users' home. These devices detect and respond to voice commands. They normally have always-on microphones, which theoretically start recording after hearing a wake word and then send the voice data to a server via internet for further processing [1].

This behavior has sparked many privacy concerns, including but not limited to what is being recorded, how the collected data is used and stored, and whether it is being protected well [4–6, 8, 9]. Much of previous work has been centered around protecting sensitive information from being leaked to adversaries [4–6]. As for what data is being collected, Amazon has made it possible for users to view, play and delete their voice data stored on its server [7]. However, our intuition is that Amazon may reveal to users only part of the voice data being collected by its smart speakers. Up until now, little has been done to verify whether Amazon is being honest with us or not—is Amazon making public exactly the audio data it collects?

In this work, we performed a series of experiments with Amazon Echo that answered the following questions: 1. Is Amazon Echo recording before

Our preliminary experiments with Amazon Echo have found that the recorded part of the same command is not consistent across different trials. This stimulates us to question and examine the behavior of Amazon Echo and dig into what is exactly being collected by it. We plan to use network traffic measurement for our work, through which previous researchers have been able to infer users' activities [6], and establish network signatures for Amazon Echo [7]. If this approach does work, we will be able to answer a whole set of research questions with it.

2 BACKGROUND

My background is here. **[[add background]]**

3 EXPERIMENT SETUP

We will begin by measuring the baseline traffic from the smart speaker in a silent environment. This should allow us to identify the regular traffic patterns of the smart speaker when it is not transmitting voice data.

Once we have baseline traffic patterns we will begin more detailed measurements. We will test pre-recorded commands of varying lengths many times, to establish correlations between recording length and traffic speed and/or duration. We will then repeat these tests in the presence of several different forms of background noise, including at least podcasts, music, and discussion among ourselves.

To determine how the speaker decides to terminate its recording, we will use several different pathological commands. We will use incomplete commands, very long commands, and commands spoken very unclearly. By correlating the traffic patterns of these commands to the more normal commands previously measured, we hope to calculate the rough length of time the recording lasts.

We will obtain whatever information we can from the cloud (e.g. through an app or API) to compare such information to our traffic pattern analysis and determine whether they match within whatever margin of error we can achieve. We will also test the behavior of the speaker in noisy environments without any commands being given to determine the rate of false recordings.

3.1 Devices

Our main device for testing will be the Amazon Echo, as it is the most popular smart speaker.

If available, we would also like to test the Apple HomePod, as Apple places a much greater emphasis on security and privacy than its competitors.

4 MEASURING TRAFFIC STREAMED TO SERVER

This measures the traffic streamed to server.

5 CONCLUSION

My conclusion is here.

6 RELATED WORK

Amazon Echo is a smart speaker developed by Amazon that is connected to the internet [3]. It has an always-on microphone that starts recording automatically after hearing certain wake word (Alexa, Echo, etc.). Voice commands following the wake word are streamed to Amazon's cloud-based intelligent command handling program, known as Alexa, for further processing. Alexa will then try to respond to users' commands. Amazon also keeps copies of the voice commands and responses, together known as response cards [7]. These response cards have been made available to users by Amazon so that they can be viewed, played, and deleted [2].

REFERENCES

[1] Amazon Echo (2nd generation) âĂŤ Alexa Speaker. https://www.amazon.com/Generation-improved-sound-powered-design/dp/B06XCM9LJ4?. (????). (Accessed on 10/15/2019).

[2] 2010. GP. (2010). https://www.amazon.com/gp/help/customer/display.html?nodeId=201602040

[3] 2019. Amazon Echo. (Oct 2019). https://en.wikipedia.org/wiki/Amazon_Echo

[4] Noah Apthorpe, Danny Yuxing Huang, Dillon Reisman, Arvind Narayanan, and Nick Feamster. 2019. Keeping the smart home private with smart (er) iot traffic shaping. *Proceedings on Privacy Enhancing Technologies* 2019, 3 (2019), 128–148.

[5] Noah Apthorpe, Dillon Reisman, and Nick Feamster. 2017. A smart home is no castle: Privacy vulnerabilities of encrypted iot traffic. *arXiv preprint arXiv:1705.06805* (2017).

[6] Noah Apthorpe, Dillon Reisman, Srikanth Sundaresan, Arvind Narayanan, and Nick Feamster. 2017. Spying on the smart home: Privacy attacks and defenses on encrypted iot traffic. *arXiv preprint arXiv:1708.05044* (2017).

[7] Marcia Ford and William Palmer. 2019. Alexa, are you listening to me? An analysis of Alexa voice service network traffic. *Personal and Ubiquitous Computing* 23, 1 (2019), 67–79.

[8] Geoffrey Fowler. 2019. Perspective | Alexa has been eavesdropping on you this whole time. (May 2019). https://www.washingtonpost.com/technology/2019/05/06/alexa-has-been-eavesdropping-you-this-whole-time/

[9] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, are you listening?: Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 102.