Alexa Privacy and Data Handling Overview

Introduction

Privacy is foundational to how we design and build Alexa and Echo devices. This white paper explains how Amazon designs Alexa and Echo devices – top to bottom – with privacy in mind.

We first provide a brief overview of Alexa systems – how audio is streamed from an Echo device to the Alexa cloud and how data is processed in the Alexa cloud. We then discuss what data from Echo devices subscribed to Alexa Smart Properties' products (e.g., Alexa for Hospitality) is and is not stored.

Alexa System Overview

This section describes what the Alexa system[1] is, and how it processes requests. We will use a simple request of "Alexa, what is the weather" to illustrate how voice requests are picked up by Echo devices, sent and processed by a series of cloud components, and finally responded to.
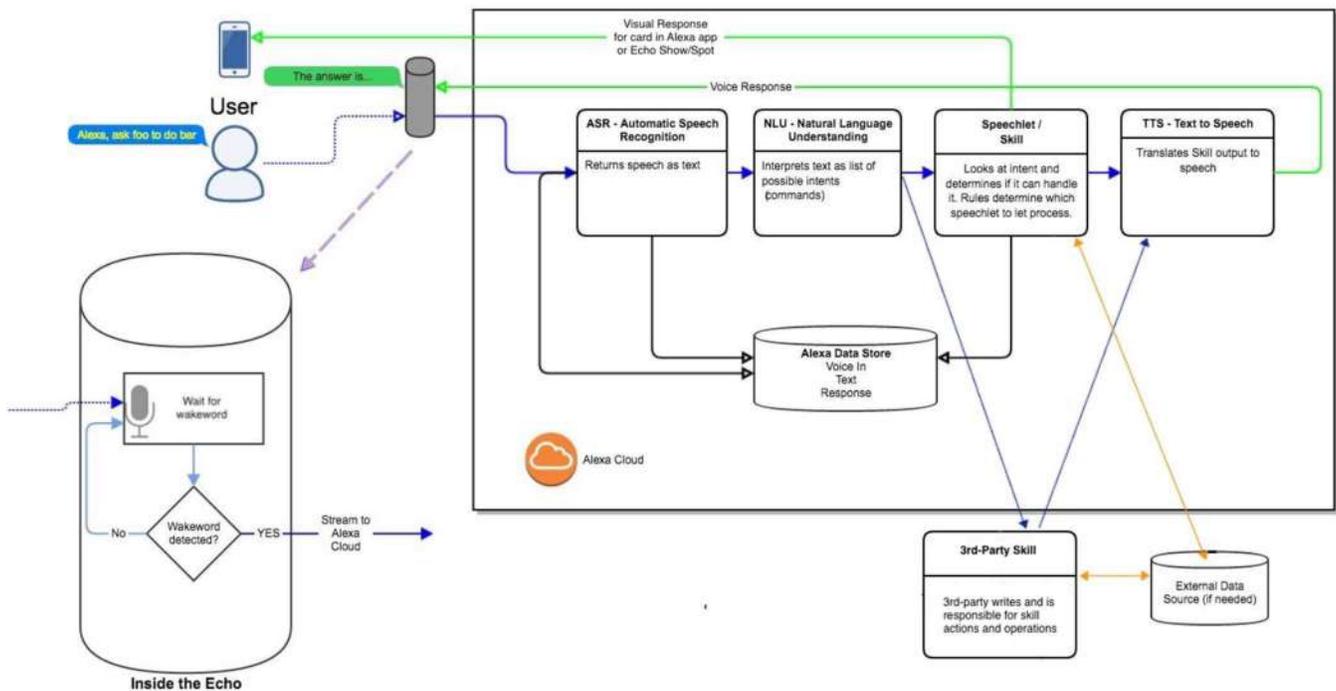

Figure: Overview of Echo and Alexa System

Echo Devices

Echo Devices and other Alexa-enabled devices are the input and output devices for Alexa. From the beginning, we have designed and built them with multiple layers of privacy protections and controls. Echo devices detects the wake word by identifying acoustic patterns that match the wake word. No audio is stored or sent to the cloud unless the device detects the wake word (or Alexa is activated by pressing a button).

---

[1] A technical overview of ASR, NLU, and Skill development is presented in Just ASK: Building an Architecture for Extensible Self-Service Spoken Language Understanding, Proceedings of the Neural Information Processing Systems 2017 Conference, March 2018.

The end-user can turn the microphone off at any time by pressing the microphone button on the device. When the button is pressed to turn the microphones off, the microphones are electrically disconnected. A dedicated red LED on the device is illuminated to indicate the microphones are off and audio is not streamed to the cloud.

With the microphone on, when the end-user says, "Alexa, (wake word detected and the Echo's light ring turns blue) what is the weather in Menlo Park?" The Echo's spins to indicate that audio is transmitted to the cloud. Data is encrypted in transit to the Cloud.

> Aside: Echo Show devices are also equipped with a camera. The blue line on screen indicates audio is streamed to the Cloud. When the Microphone/Camera Off button is pressed, a dedicated red light is illuminated and the (Ø) symbol is shown on screen. The button also disconnects power to the camera and the microphones.

Automatic Speech Recognition (ASR)

The first system in the Alexa cloud to receive data is Automatic Speech Recognition (ASR). It takes the audio stream and turns it into a text string (or set of possible text strings) that are sent to the Natural Language Understanding (NLU) system. In our case, possible strings could be:

- "what is the weather in menlo park"
- "watt is the weather in menlo park"
- "what is the whether in menlopark"
- "watt is the whether in men lo park"

These strings, and their related confidence scores, are used to improve speech recognition. The string with the highest score (i.e., the one that Alexa acts on) is stored.

Natural Language Understanding (NLU)

The "Natural Language Understanding" (NLU) interprets the recognition result and produces an intent. The NLU performs:

- intent classification (determining in this case that the user wishes to get the weather, and returning a Weather intent)
- entity recognition (determining that the user requested the location "Menlo Park")
- slot resolution (determining the location identifier for the location "Menlo Park" that can later be used to retrieve the correct weather for that location)

The service now looks at the intent as "Weather" and routes the request to the proper application (Skill) with the slots filled in for location (94025) and time (today).

Skills

Skills extends what the Alexa system can do. In our example, the skill takes the input (location [Menlo Park] and time [today]) and retrieves the appropriate information from the designated data source. It then turns raw data (in our example, temperatures and forecast) into a textual response, formatted with Simple Speech Markup Language (SSML). The response is sent to the response system.

When an end-user uses a skill, we may exchange related information with the developer of that skill, such as answers to trivia questions, ZIP code of a weather request, or the content of requests (but not voice recordings).

Alexa Smart Properties supports skills published to the Alexa Skills Store and certified property-specific skills hidden from the Store. The following types of skills are not supported: child-directed skills, skills that request personally identifiable information (PII), skills that contain mature themes or content, and skills that enable shopping.

Responses – Text to Speech (TTS)

The response system takes the SSML that is produced by the skill, uses TTS to generate the audio speech file, and streams the audio to the appropriate device. For many skills, this ends the interaction. Other skills are interactive and will ask follow-on questions that require answers. Echo devices show the blue light ring when audio is being recorded and streamed to the cloud, including when waiting for a response.

Data Retention

To strengthen end-user privacy, voice recordings from devices subscribed to Alexa Smart Properties' Products are not saved.

Aggregated metrics that are not connected to individual end-users — Alexa requests and dialogue count at the property and organization level, and usage of device by intent category level (e.g., music) — are stored in databases. Access to metrics is restricted to teams and individuals that need this data to perform their work.

Appendix: Privacy & Security FAQs

1. How do Echo devices recognize the wake word?

By default, Echo devices are designed to detect only your chosen wake word (Alexa, Amazon, Computer, or Echo). The device detects the wake word by identifying acoustic patterns that match the wake word. Once the wake word is detected, the device begins recording audio to send to the cloud, including a fraction of a second of audio before the wake word. No audio is stored or sent to the cloud unless the device detects the wake word (or Alexa is activated by pressing a button).

2. Can I turn off the microphone on Echo devices?

Yes. Echo devices are equipped with a microphone off button. When the button is pressed, the power to the microphones is disconnected and a dedicated red light is illuminated. When the microphones are turned off, your device cannot record and stream audio to the cloud, even when you say your chosen wake word.

3. How do I know when an Echo device is streaming my voice to the Cloud?

When Echo devices detect the wake word or when the Action button available on some Echo devices is pressed to activate Alexa, a visual indicator appears on the device to indicate that the device is recording your request to stream to the cloud. For instance, a light ring on the Echo will turn blue or a blue bar will appear on Echo Show. When you use the wake word, the audio stream includes a fraction of a second of audio before the wake word, and closes once your request has been processed. Organizations, managing shared Echo devices using Alexa Smart Properties' Products, can also configure Echo devices to play a short audible tone any time audio is sent to the cloud.

4. Are voice recordings saved?

No. Voice recordings from devices subscribed to Alexa Smart Properties' Products are not saved.

5. Can I turn off the camera on Echo devices with a camera?

Yes. You can turn off the camera on Echo devices with a camera by pushing the Microphone/Camera Off button on your device. When the Microphone/Camera Off button is pressed, the power to the camera and the microphones is disconnected and a dedicated red light is illuminated. Some Echo devices with a camera, like Echo Show 5, have an additional built-in camera shutter. When the camera is turned off by pressing the Microphone/Camera off button or the camera shutter is covering the camera, you will not be able to use any features that rely on the camera.

6. Are audio calls recorded?

Audio between Alexa devices are encrypted end-to-end. Amazon does not store your audio calls, or information about the content of those calls. If you use Alexa to make a call to a landline or mobile phone, the part of that call that takes place over the telephone network is not encrypted.

7. When an organization manages shared devices using Alexa Smart Properties' Products, what information does that organization have access to?

The organization can see and control which skills are enabled on a shared device, the property unit where it's assigned, and the settings applied to the device. The organization has access to aggregated engagement and device usage metrics (unique device count, active device count, top 10 dialogue categories or domains by dialogue count, and top 10 skills by dialogue count) at the organizational and property level. The organization also has access to intent level metrics (dialogue distribution per intent) and daily dialogue count for their custom hidden skills.

The organization has no access to any voice recordings or transcripts of what end-users said to Alexa. Voice recordings from shared devices subscribed to Alexa Smart Properties' products are not saved.

8. What data do skill developers have access to when their skills are enabled on devices subscribed to Alexa Smart Properties' products?

Skill developers can view skill metrics provided in the Alexa Skills Kit developer portal. They also have access to any logs written from their back end code.