Alexa Confidentiality and Data Handling Overview

Getting customer confidentiality and privacy right takes careful attention, and Amazon has designed and built privacy, confidentiality, and security deeply into Echo hardware, the Alexa service, and the Alexa services used by 3rd-parties to build their own Alexa-capable hardware. Inherent in confidentiality and privacy is customer control. This paper provides background on how Alexa works, and gives details on customer control: both control over when audio is streamed from an Alexa-capable device to the Alexa cloud, and control over their data and how it used while in the Alexa cloud.

Information is presented in three sections. First is an end-to-end overview of the Alexa system. Next is a description of how devices detect the Alexa wake word and begin streaming audio to the cloud. Last is a description of data use, management, and retention within the Alexa cloud.

What is the Alexa System?

We must first understand what the Alexa system is, and how it processes requests. For purposes of this discussion, the Alexa system is the system in the cloud that understands speech and carries out or responds to customer interactions. It does so through several components, which have the majority of the "smarts": Automatic Speech Recognition, Natural Language Understanding, and Response. Some responses are provided by third party services through "skills." The 3rd-parties that write and publish those skills are responsible for their skill's behavior.

Customers interact with Alexa via an Alexa-enabled device. These devices are made by Amazon—including the Echo and other Alexa-enabled devices such as some FireTVs, and Fire Tablets—as well as by 3rd-parties, who can use the Amazon-provided Alexa Voice Service (AVS) software development kit (SDK) to build their own devices or integrate Alexa into existing devices.

To start, we begin with a very simple tour through the entire system, demonstrated with the example request of "Alexa, what is the weather in Miami," so we can see how the request is picked up by an Alexa device, sent through voice recognition, interpreted, acted upon, and then responded to.

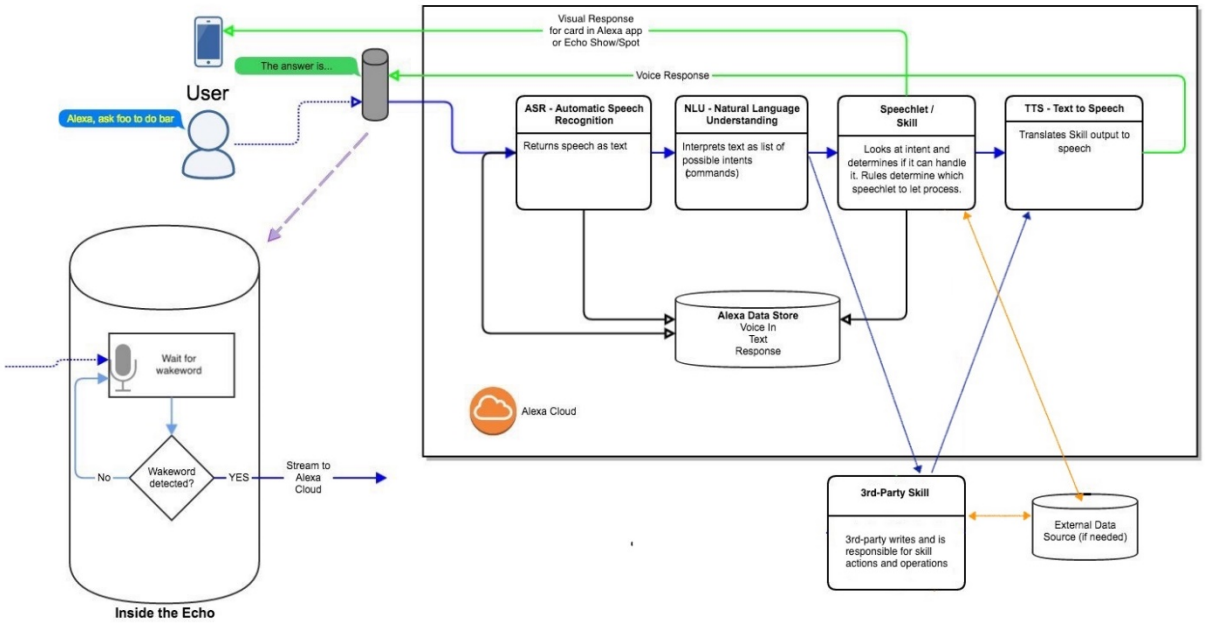

Figure: Overview of Echo and the Alexa System

ACTIVATION

An interaction with Alexa begins with activation of the Alexa-enabled hardware, either via a button press or by device detection of the wake word. Wake word detection is discussed in detail below. Once activated by the wake word or the action button, the device opens an audio stream to the cloud and sends the request to Alexa to respond accordingly. Alexa will end the stream immediately once the user ends the conversation or if Alexa detects silence or speech that isn't intended for Alexa, a light or similar indicator on the device indicates when audio is being streamed to the cloud. The system is designed so that communication between the Alexa endpoint and the Alexa cloud is encrypted and protected using TLS 1.2.

AUTOMATIC SPEECH RECOGNITION (ASR)

Automatic Speech Recognition (ASR) takes the audio stream and transcribes it (i.e. turns it into a text string or set of possible text strings). Transcriptions are sent to the Natural Language Understanding (NLU) system.

In our case, examples of possible transcriptions could be:

- "what is the weather in miami"
- "watt is the weather in miami"
- "what is the whether in miami"
- "watt is the whether in my amy"

These transcriptions, and their related confidence scores, are used to improve speech recognition. The transcription with the highest score (i.e., the one that Alexa acts on) is also stored. Users can view this data in the Alexa companion app or on the Alexa website. This information is not available to Alexa for Business users or administrators. However, if a user wants to know what Alexa heard from a device in the previous minute or so, they can say "Alexa, what did I just say?" and Alexa will repeat back what it thinks that it heard. A fraction of one percent of the audio recordings are manually reviewed to ensure that the transcription is correct, and that it is being properly understood (next step).

NATURAL LANGUAGE UNDERSTANDING (NLU)

The "Natural Language Understanding" (NLU) interprets the transcription result and produces an intent—an instruction to the Alexa system to tell it what to respond to. The NLU system 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 "Miami"), and
- slot resolution (determining the location identifier for the location "Miami" 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 (Miami) and time (today). In this case, the Alexa system needs to access an external data source to get the relevant temperature and weather conditions.

The data about the chosen intent, and information related to entity recognition and slot resolution, is stored for machine learning purposes.

SKILLS

Skills are like "apps" for Alexa and extend what the Alexa system can do. Amazon has designed our skills program to share only limited information with the third-party developers of those skills. For example, voice recordings are not shared with skills.

In the above example, the skill takes the input (location: Miami and time: today) and retrieves the appropriate information from the designated data source, which returns the needed data. The skill then formulates its response, taking the raw data (in this case, temperatures as well as the forecast) and constructs a textual response formatted with SSML (simple speech markup language) which tells the next step, TTS, how to respond. Once the response is generated, it is sent to the response system.

RESPONSES (TTS)

The response system takes the SSML that is produced by the skill, uses text-to-speech (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 are designed so that the blue ring on the Echo device is lit when the device is waiting for a response to a question that Alexa has asked.

The text of the response is stored by the Alexa system so that users of personal devices can review past answers using the Alexa app. Access to this data is not available to Alexa for Business users or administrators for devices managed by Alexa for Business. In addition, the response can be used by the Amazon team who built the specific skill to ensure that Alexa is providing relevant answers to queries and that the TTS system is properly translating the text to speech.

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 (https://arxiv.org/pdf/1711.00549.pdf)

When is Audio Streamed to the Cloud?

Customer interactions with Alexa start with an endpoint. The endpoint is responsible for detecting when it has been activated, sending an audio stream to the cloud, and receiving and playing the response. There are many different types of endpoints, but they can be broadly categorized by activation method: either wake word detection or action button. Those devices that use wake word detection can further be categorized by the technology they use to detect the wake word. Regardless of the method or technology used to activate Alexa, the process in the cloud described above is the same for all Alexa endpoints. In addition, Echo and other Amazon built devices are designed from the beginning with multiple layers of security and privacy protections and controls. Devices built by 3rd-parties must meet the security requirements that Amazon has established for 3rd-party vendors[1].

The system is designed so that communication between the Alexa endpoint and the Alexa cloud is encrypted and protected using TLS 1.2.

ACTIVATING WITH THE ACTION BUTTON

The simplest form of Alexa activation uses an action button on the device (or remote connected to the device). For devices that activate only with an action button, the microphone is off and not receiving any

[1] https://developer.amazon.com/docs/alexa-voice-service/security-best-practices.html

signals until activated. Action buttons can take two forms: push-to-talk, where the microphone is active only when the button is pressed and tap-to-talk, where the microphone turns on with the press and remains on until the user interaction with Alexa completes. While the microphone is active and the Alexa system is processing the request, the user is notified that streaming is occurring by a visual cue (such as a light), an audible cue, or both. When the interaction is complete, no audio is processed by the device and sent to the Alexa cloud. Some devices use both a wake word and also have an action button. For those devices, when the interaction is complete, the device goes back to its wake word detection mode.

ACTIVATING WITH A WAKE WORD

The other way that Alexa is activated is through the use of a wake word. In this case, devices use on-device technology to detect when the wake word is spoken and then turn on the audio stream to the Alexa system in the cloud. As with an action button interaction, when the Alexa system starts processing the interaction, a visual or audible indicator will make the user aware that audio is streaming to the Alexa system. When the interaction is complete, the visual indicator turns off or audible indicator sounds to signify the end of the interaction.

There are different methods of wake word detection, depending on the type of device. Amazon-built devices, such as those in the Amazon Echo family, use on-device keyword spotting designed to detect when a customer says the wake word. This technology inspects acoustic patterns in the room to detect when the wake word has been spoken using a short, on-device buffer that is continuously overwritten. This on-device buffer exists in temporary memory (RAM); audio is not recorded to any on-device storage.[2] The device does not stream audio to the cloud until the wake word is detected. If the device does not detect what it thinks is the wake word no audio is sent to the cloud.

When devices are built by 3rd-parties we make available, but do not require 3rd parties to use, our Amazon-developed wake word engine. In any case the 3rd-party developers must comply with the Alexa Voice Service (AVS) guidelines and requirements and have their devices reviewed and approved by Amazon before they can be marketed to customers as "Alexa Built-In" products. Details on specific implementation details for 3rd-party devices can be obtained from the manufacturer. Information on the design and compliance guidelines and requirements for 3rd-party devices can be found at the https://developer.amazon.com/alexa-voice-service/launch website.

USER CONTROL

With either an Amazon or 3rd-party developed Alexa endpoint, the customer is in full control and can easily disable the device's microphone. When the microphone disable button is pressed on Amazon-built Echo devices with cameras, when the microphone button is pressed the camera is also disabled. For devices that only use an action button, no action is required, since the microphone is only active based on a user press. For devices that activate using a wake word, there must be an always-available control to turn off the Alexa wake word or disable the device's its microphones. The customer can turn the microphones off at any time using this control. For Amazon-built devices, when the button is pressed to turn the microphones off, the microphones are electrically disconnected and a dedicated red LED on the microphone button is illuminated to indicate the microphones are off.[3] As a safeguard, Amazon designed the circuitry of Echo smart speaker

[2] Some Echo devices can do limited on-device voice processing in order to provide local functionality when a connection to the internet (and thus Alexa) is not available. This enables users to do things like control smart home devices and set alarms and reminders even without a network connection. In these cases, when there is no network connection, these commands are stored locally so they can be sent to Alexa once the network connection is reestablished.

[3] Certain Alexa-enabled devices intended for use by a single person, such as the Echo Loop (ring), Echo Frames, and Echo Buds, which also communicate via a paired mobile phone, have controls and indicators located in the phone-based Alexa companion app.

devices so that power can only be provided either to this dedicated red LED or to the device microphones, not to both at the same time. As a result, if the dedicated red LED is illuminated, the microphones are off and cannot record and stream audio to the cloud.

INTERACTIONS BETWEEN THE ALEXA ENDPOINT AND THE ALEXA CLOUD

As described above, when the wake word is detected or the action button is pressed a connection to the cloud is opened up and audio begins streaming. With devices using wake word detection the audio stream consists of a fraction of a second of audio prior to the wake word and continuing until the Alexa system in the cloud turns off the audio stream.

If Alexa is activated using the wake word, the first step that occurs when the stream reaches the cloud is that the audio is reanalyzed using the more powerful processing capabilities of the cloud to verify the wake word was spoken. These additional algorithms are in the cloud, and not on the device, for reasons including requiring more processing power than the Alexa device has available or using machine-learning derived models based on recent learnings. The on-device algorithms are automatically updated on a regular basis. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio.If the wake word is verified (or if Alexa was activated using the action button), the ASR and NLU systems process the customer's request so Alexa can respond appropriately. As Amazon's speech recognition system analyzes the audio stream, the system continually attempts to determine when the customer's request to Alexa has ended and then immediately ends the audio stream.

For most devices, an indicator light then typically flashes blue/light blue until the response is ready for playback. It then sends the response (the blue light pulses while Alexa is speaking), and the Alexa device returns to monitoring for its wake word.

HOW LONG DOES THE ALEXA DEVICE STREAM AUDIO TO THE CLOUD?

As Alexa analyzes the audio stream, the service continually attempts to determine when the customer's request to Alexa has ended and then immediately ends the audio stream. In some circumstances, in response to customer commands, the stream will open again for a customer to follow up, including if the customer's request involves multiple interactions. For example, if you say, "Alexa, set the timer," Alexa will respond with "Timer for how long?" and will open the audio stream to wait for your response. Similarly, an interactive skill like "20 Questions" will ask questions and Alexa will open the audio stream for your response.

Customers may also elect to enable the Follow-Up mode setting (on a device-by-device basis). Follow-Up mode allows Alexa to respond to a series of requests in rapid succession without the customer needing to repeat the wake word for each request, but only after being woken by an initial request with the wake word.

In all cases, a visual or audible cue will indicate to customers that the Alexa device is streaming audio to the cloud. Customers can also enable an audible tone that plays at the start and end of each request.

ALEXA CALLING

When using Alexa calling, the light ring (or bar) on Amazon-built devices will glow green, to indicate that the microphone is on and audio is streaming. In this case, the audio is not being streamed to the Alexa ASR and NLU systems. Instead, it is being routed to either another Alexa-enabled product (for Alexa-to-Alexa calling) or to the phone system (if placing a telephone call). Calls are not recorded by Alexa. During the call, the Alexa device is still monitoring for its wake word, so that you can say things like "Alexa, end call". Even during a call, you will note that there is a visual or audible cue to indicate when its wake word is detected, to show that the audio directly after the wake word is going to Alexa.

SUMMARY OF DEVICE ACTIVATION AND AUDIO STREAMING TO ALEXA

To wrap up this section, we have addressed the following:

1. Alexa devices are the input/output devices for the Alexa system, which is in the cloud.
2. Audio is sent from an Alexa endpoint to Alexa in the cloud *only* when the wake word is detected or the action button is pressed. Otherwise, no audio is sent to the cloud.
3. A visual or audible cue indicates when audio is being sent from an Alexa endpoint to the Alexa system in the cloud
4. Alexa double-checks, with cloud-side verification, that the wake word was really spoken.
5. Alexa ends the audio stream when it determines that the customer is no longer talking to Alexa.
6. Alexa wake word detection is under the user's control (either via an on-device microphone button or some other control).
7. On Echo smart speaker devices, the circuitry of the red LED on the microphone button is electronically connected to the microphones. If the red LED on the microphone button is lit, the microphones are off.
8. During an interaction with Alexa, there will be a visual or audible indication if Alexa is waiting for input.
9. Alexa devices built by 3rd-parties have features similar to those provided by Amazon-built devices, and must comply with Amazon policies and be approved by Amazon before marketed to customers as "Alexa Built-In" devices.

Data Retention and Use in Alexa

Alexa is designed to get smarter every day—this is accomplished through the power of machine learning and the cloud**.** When a customer says the wake word, their subsequent phrases are processed and stored in the cloud to respond to the customer's request and to improve the customer's experience and Amazon's services, including training speech recognition and natural language understanding systems so Alexa can better understand customers' requests. Amazon gives customers control over their data in many ways—they can delete their voice recordings at any time and can control how the data is used.

Data Use

Different types of data are used and stored by the Alexa system to provide the Alexa service. Configuration parameters are set by the user either on the device or using the Alexa app. These parameters include such things as the device location (set by the administrator or user), preferred time zone and unit measures, volume level, and other preferences.

Amazon uses your requests to Alexa to train our speech recognition and natural language understanding systems using machine learning. Training Alexa with real world requests from a diverse range of customers is necessary for Alexa to respond properly to the variation in our customers' speech patterns, dialects, accents, and vocabulary and the acoustic environments where customers use Alexa. This training relies in part on supervised machine learning, an industry-standard practice where humans review an extremely small sample of requests to help Alexa understand the correct interpretation of a request and provide the appropriate response in the future. For example, a human reviewing a customer's request for the weather in Austin can identify that Alexa misinterpreted it as a request for the weather in Boston. Our supervised learning process includes multiple safeguards to protect customer privacy. We limit the information available to individuals

who are assigned transcription and annotation tasks, and we use multi-factor authentication to restrict access, service encryption, and audits of our control environment to protect it.

Amazon also uses requests to help personalize responses. For example, Alexa will learn what kind of music you like so that a request like "Alexa, play some music" will all produce results that reflect recent requests or preferences.

Data Retention

Data is stored in multiple forms and for multiple purposes in various Amazon services, such as S3 and DynamoDB (under the control of the Alexa service). Data is retained to allow Amazon to provide the service to customers (including allowing enrolled users to review and play back their voice recordings) and to build, test, debug, and improve our systems.

Only those who have an approved need to access certain data to accomplish their job are given access to that data—access is granted via specific, audited permissions and access to customer data requires review and approval by the responsible managers. Additionally, managers must confirm, on a quarterly basis, that individuals should be members of teams that have access to this data.

Sensitive customer data in the Alexa system (such as voice recordings) is stored in databases and encrypted at rest and in transit, using Amazon's internal key management systems.

Some system level data is also stored in log files, for either service troubleshooting purposes, or security incident resolution. Troubleshooting logs contain information necessary for developers to troubleshoot the Alexa system, but do not contain customer voice recordings or data derived from customer voice recordings, such as slot values or the TTS response. Access to these logs is restricted to teams needing access to this data to perform their business functions. Troubleshooting logs are encrypted and their access is audited.

Security logs are retained for purposes of audits and are restricted to those operating in security incident roles. They contain data that describe (1) when systems or users authenticated themselves to the system and (2) which systems and users accessed which data, and when. Again, these logs are encrypted and the data in them is used to ensure that system use complies with applicable policies.

Amazon applies retention policies to data to minimize the data we retain. Data is retained when it serves a business purpose (including providing the service to customers and improving our systems) or as necessary to comply with law.

Metrics are stored in databases. Metrics are used for internal business processes, to direct system improvements, for systems performance analysis and reporting, and for customer reports. Access to metrics is restricted to the teams and individuals that need this data to perform their work. As with other data access, these permissions are reviewed and approved at least quarterly.

Data Use By 3rd-Party Skills

Customer's personal information (e.g. name, address) are not released to the 3rd-party unless specifically requested to be shared by the customer. We also use a permission framework similar to the one used by mobile devices, which requires customers to grant permission to share certain data with skill developers–e.g., Lyft could request permission to access the address the customer has set for their Echo device so Lyft can send a ride to that location, and we would only share that address with Lyft after the customer granted permission. Even when a customer links their Amazon account to a 3rd-party skill account (e.g., when a customer links their account with Lyft, so that Lyft always knows which Lyft account to charge rides to), the

3rd-party doesn't receive the customer's Amazon account identifiers. Instead, they receive a token. Amazon can identify an Amazon customer from this token, but skill developers cannot. Each time a customer talks to a skill, the skill gets the same token for that user. However, different skills get different tokens when they talk to the same user, so even if skill developers share data, they cannot determine that they share common customers based on data that Amazon has shared. Each 3rd-party skill has their own policies concerning storage and retention of skill-specific data. Customers can read the 3rd-party skill developer privacy policies, and see what customer information is being asked for, on the skill detail page on the Amazon website.

User Control of Retained Data

Amazon gives customers multiple ways to manage their data. As stated above, audio recordings are used to improve Amazon services. The tools for managing the audio are slightly different for customers using personal devices registered to their own Amazon account (including enrolled users who have linked their personal accounts to their company's Alexa for Business account) and customers who have devices that are managed directly by Alexa for Business (shared devices).

CLEARING AUDIO RECORDINGS

For personal devices, customers can review voice recordings associated with their account and delete those voice recordings one by one or all at once by visiting *Settings* > *Alexa Privacy* in the Alexa app or https://www.amazon.com/alexaprivacysettings. They can delete those voice recordings all at once for each of their Alexa-enabled products by visiting Manage Your Content and Devices (https://amazon.com/mycd). Customers can also delete their recordings by voice. Once enabled, they can delete the voice recording of their last request by saying "Alexa, delete what I just said" or can delete all the voice recordings for the day by saying "Alexa, delete everything I said today."

Alexa for Business enterprise customers, those with shared devices, do not have access to review audio recordings, but can still manage the audio recordings made on devices they manage. Customers using shared devices can use the same voice commands described above (though with shared devices, the "delete everything I said today" command clears all audio recordings made by the specific device receiving the command, and not to all utterances made by that customer to all devices).

Alexa for Business administrators have two additional ways to manage the audio recordings in the Alexa system. First, there is a button next to each device entry in the Alexa for Business console that allows the recording history for that device to be cleared at any time. Second, there is an API available (*DeleteDeviceUsageData*) that will clear the previous history of voice input data for the device. Enterprises can use the API to clear device history on a regular basis.

For all users, when a voice recording is deleted, Amazon deletes the text transcript associated with the customer's account of the customer's request.

CONTROLLING USE OF DATA

Customers can control how voice recordings that are retained by Amazon are used. As described earlier, Amazon makes use of this data to improve algorithms and make Alexa smarter. However, customers have the option to limit the use of voice recordings so they are not used to develop new features or manually reviewed as part of our supervised machine learning process to help improve Amazon services. Individuals can do so from the Alexa Privacy Center and Alexa for Business users can do so from the Alexa for Business Console.

Summary of Data Retention and Use

To wrap up this section, we have addressed the following:

1. Voice recordings from Alexa are used to improve the Alexa system through supervised and non-supervised machine learning.
2. Customers have control over their voice data and how it is used by Amazon.
3. Enterprise customers can use voice, console, and APIs to clear voice data history.
4. When a voice recording is deleted, the related text transcript associated with the customer's account is also deleted.
5. Customers can instruct Amazon not to use their voice recordings to develop new features and not to manually review them as part of our supervised machine learning process to help improve our services.