Alexa, Who Has Access to My Data?

Amazon Reveals Private Voice Data Files

At some point, we simply forget all our fears and end up inviting Alexa to share the bedroom or the shower. What's wrong with having a chat with Amazon's virtual assistant in the privacy of our own homes? Things only start to go bad when Amazon gets sloppy and sends your voice files to a stranger. Recently, the critics' warnings came true at amazon.de.

By Holger Bleich

Sales of virtual assistants for home use are booming. Even though Amazon doesn't publish sales figures, we know that Echo products are out there in millions of homes. This Christmas, hundreds of thousands more will find their way into our lives.

Amazon always confronts data protection concerns with the same arguments: Users know what Echo hears and when. According to the terms and conditions every user agrees to, only the user and Amazon can listen to what the device records. This promise is the cornerstone of the trust that Amazon demands from Alexa users.

A company like Amazon would be crazy to treat sensitive personal data with anything but the utmost care. However, a recent case researched by c't shows that the processes used to handle personal data at amazon.de have serious issues. This is the "worst case scenario" that consumer and data protection activists have been warning us about.

Contaminated Data

Indirectly, the case has to do with the consequences of the rights enshrined in the new EU General Data Protection Regulation (GDPR). These rules grant consumers extensive rights when it comes to reviewing the personal data collected by companies like Amazon. In August 2018, amazon.de customer Martin Schneider (not his real name) used these rights to ask Amazon to release the personal data it has on file about him. A couple of months later, Schneider was sent a download link to a 100MB ZIP file.

About 50 of the zipped files contained data relating to everyday things like Amazon searches, but there were also around 1,700 WAV files and a PDF cataloging unsorted transcripts of Alexa's interpretations of his voice commands. Schneider was extremely surprised to find these files as he doesn't use Alexa and doesn't own any Alexa-enabled devices. He listened to some random sample files but didn't recognize any of the voices they contained.

On November 8th, Schneider mailed customer service at amazon.de to tell them the files he had received were obviously someone else's. He also asked for more information about who they might belong to. He never received a reply but, shortly after, found that the download link to the files was dead. Schneider had already saved the files locally and contacted c't in mid-November as he was worried about the implications of what he had found. He was also concerned because Amazon had gone silent and he felt that the victim should be informed about the leak.

Infringement of Privacy

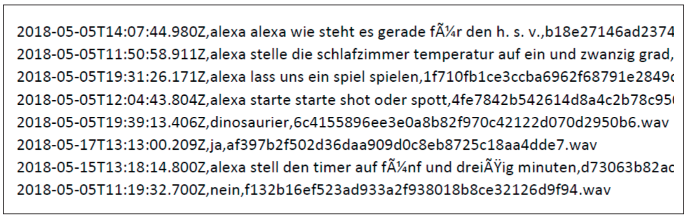
We asked him to send us some of the files (confidentially of course) so that we could get an idea of what they contained. They enabled us to piece together a detailed picture of the customer concerned and his personal habits. It was obvious that 'Customer X' uses Alexa in multiple locations. He has at least one Echo at home and has a voice-controlled Fire box connected to his TV. A female voice also spoke to Alexa, so there was clearly a woman around at least some of the time.

Alexa was obviously able to hear our 'subject' in the shower, and commands given to thermostats and the like showed that he uses Alexa to control various smart home appliances. He uses Alexa at home, on his smartphone, and when he is out and about. The recordings we received covered the entire month of May.

We were able to navigate around a complete stranger's private life without his knowledge, and the immoral, almost voyeuristic nature of what we were doing got our hair standing on end. The alarms, Spotify commands, and public transport inquiries included in the data revealed a lot about the victims' personal habits, their jobs, and their taste in music.

Using these files, it was fairly easy to identify the person involved and his female companion. Weather queries, first names, and even someone's last name enabled us to quickly zero in on his circle of friends. Public data from Facebook and Twitter rounded out the picture.

We couldn't find a phone number, so we used Twitter to ask the victim to contact us. He called back immediately and we explained how we found him. We had scored a direct hit and Neil Schmidt (not his real name) was audibly shocked when we told him about the personal data Amazon had sent to a stranger. He started going through everything he and his friends had asked Alexa and wondered what secrets they might have revealed. He also confirmed that we had correctly identified his girlfriend.



```
2018-05-05T14:07:44.980Z,alexa alexa wie steht es gerade fÃ¼r den h. s. v.,b18e27146ad2374
2018-05-05T11:50:58.911Z,alexa stelle die schlafzimmer temperatur auf ein und zwanzig grad,
2018-05-05T19:31:26.171Z,alexa lass uns ein spiel spielen,1f710fb1ce3ccba6962f68791e2849
2018-05-05T12:04:43.804Z,alexa starte starte shot oder spott,4fe7842b542614d8a4c2b78c95
2018-05-05T19:39:13.406Z,dinosaurier,6c4155896ee3e0a8b82f970c42122d070d2950b6.wav
2018-05-17T13:13:00.209Z,ja,af397b2f502d36daa909d0c8eb8725c18aa4dde7.wav
2018-05-15T13:18:14.800Z,alexa stell den timer auf fÃ¼nf und dreiÃŸig minuten,d73063b82ac
2018-05-05T11:19:32.700Z,nein,f132b16ef523ad933a2f938018b8ce32126d9f94.wav
```

The transcripts reveal a lot of personal data

"Unfortunate Mishap"

The fact that Amazon linked a customer's data to the wrong person and didn't notice the mistake points to a severe lack of control over the processes involved. It is obvious that no serious checks took place. The situation is worsened by the fact that Martin Schneider received no reply when he informed Amazon of the error. Furthermore, according to the victim (Neil Schmidt) Amazon didn't contact him either.

Amazon's data protection systems are obviously flawed on multiple levels. We contacted Amazon about the case without letting on that we had identified the victim. According to the law, Amazon is obliged to contact the data protection authorities within 72 hours of discovering such a breach, and we wanted to find out if they had actually done this (see the interview below).

Amazon declined to answer our questions, but wrote to us a few days later saying that the case in question was an "unfortunate mishap" that was the result of human error. They also explained that they had resolved the issue with both customers and had introduced measures to improve the processes involved.

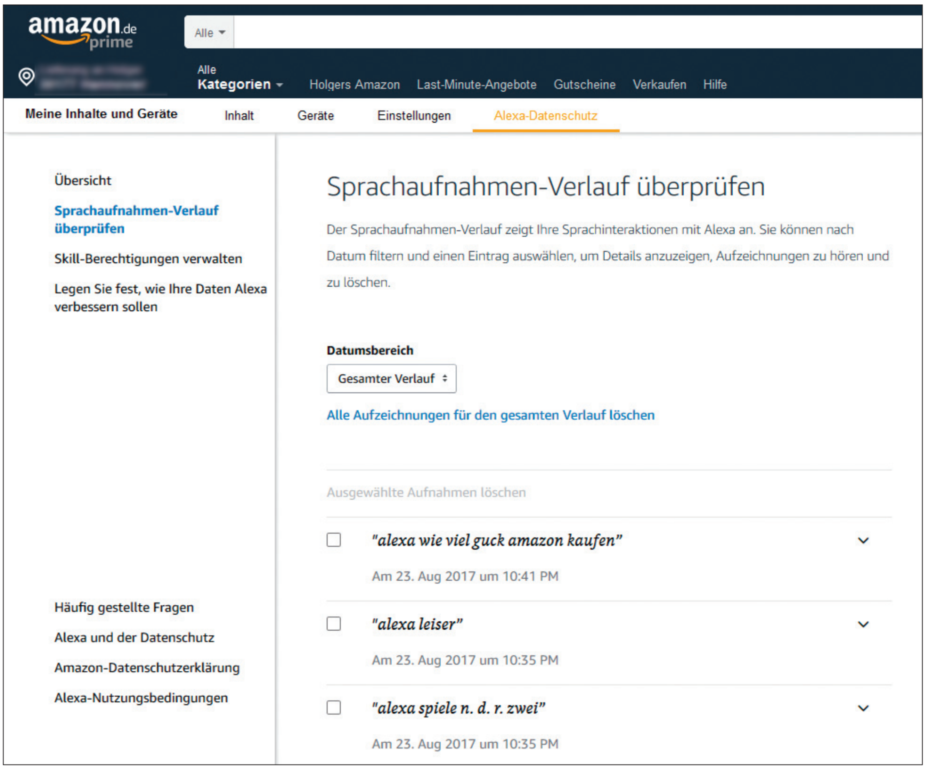
According to Martin Schneider, Amazon "resolved" the issue by having someone call him (three days after we contacted Amazon) to explain that one of their staff had made a one-time error. On the same day, Amazon also called Neil Schmidt to tell him that his Alexa voice files had fallen into the wrong hands. He had filed a GDPR information request, and customer support told him that some of his data had been released to the wrong customer as a result. Amazon also claimed that they had discovered the error themselves.

Amazon's Endless Appetite for Data

This data privacy disaster would never have occurred if Amazon had deleted the voice files in a timely fashion instead of saving them indefinitely in the cloud. However, Amazon's data privacy FAQ says it saves the files to help it develop its voice and language recognition systems. Amazon also stresses the fact that customers can review and delete voice recordings at amazon.de/alexaprivacy. It is not clear how many customers know about this option or how many actually make use of it.

This policy is problematic from a data privacy point of view. We contacted GDPR expert Dr. Carlo Piltz to help us analyze the case. He tells us that the GDPR requires default settings that explicitly protect the user's privacy, although this doesn't prohibit companies from collecting and processing personal data for appropriate purposes. In our case, Amazon's stated aim of offering a functional, AI-based voice recognition service that adheres to the principles of data economy justifies the means. The crux of the matter is deciding which data are required and how long they need to be saved to fulfill the specified purpose.

It remains unclear whether the case in question was an isolated incident, and Amazon still hasn't told us whether it fulfilled its legal obligation to inform the appropriate authorities. None of this really helps Neil Schmidt, who can count himself lucky that his voice files ended up in our hands and not posted on social media. By way of compensation, Amazon gave Schmidt a free Prime membership as well as new Echo Dot and Spot devices! *hob@ct.de*



Sprachaufnahmen-Verlauf überprüfen

Der Sprachaufnahmen-Verlauf zeigt Ihre Sprachinteraktionen mit Alexa an. Sie können nach Datum filtern und einen Eintrag auswählen, um Details anzuzeigen, Aufzeichnungen zu hören und zu löschen.

Datumsbereich: Gesamter Verlauf

Alle Aufzeichnungen für den gesamten Verlauf löschen

Ausgewählte Aufnahmen löschen

☐ *"alexa wie viel guck amazon kaufen"*
Am 23. Aug 2017 um 10:41 PM

☐ *"alexa leiser"*
Am 23. Aug 2017 um 10:35 PM

☐ *"alexa spiele n. d. r. zwei"*
Am 23. Aug 2017 um 10:35 PM

Amazon saves Alexa voice recordings indefinitely in the cloud, although they can be deleted manually via the customer's account profile

Significant Risk

Attorney Dr. Carlo Piltz advises large and medium-sized companies on national and international privacy law, and hosts the data security blog delegedata.de. He says that Amazon should definitely have informed the victim about the disclosure of his personal data.

c’t: Should Amazon have informed the victim as soon as they found out about the data leak?

Carlo Piltz: Companies are legally obliged to inform the victim and the authorities when a data privacy breach occurs with a normal or significant level of risk. However, the GDPR stipulates different information requirements depending on who is being informed.

The responsible party (in this case, Amazon) is obliged to report to the regulatory authority within 72 hours of discovery of a data privacy breach. The victim only has to be informed immediately (i.e., without unnecessary delay) if the risks involved are significant. In other words, there is no specific time constraint for informing the victim. What “without unnecessary delay” means in the context of the GDPR is currently open to interpretation. In the context of German civil law, this generally means within 14 days. However, the GDPR is a pan-European law and is subject to differing interpretations from country to country.

c’t: Is Amazon’s behavior punishable?

Piltz: Disclosure of private audio files to an unauthorized third party contravenes various stipulations of the GDPR, including the requirement to maintain reasonable data security and the obligation to ensure that data is only released to its authorized owner. The regulatory authority has to decide on the appropriateness and scope of any fines it levies. The GDPR includes a sliding scale of fines that can be applied from case to case, depending on the details involved. Was the data leak deliberate or negligent? Did the responsible party cooperate fully with the authorities and attempt to limit the effects of the leak? There is a broad range of factors that affect each case, and we are still learning how to interpret and apply them. In this case, it is now up to the authorities to investigate and decide whether to fine Amazon.