ON THE PLATFORM EPISODE: CAN YOU TRUST YOUR SMART SPEAKER?

VIDEO TRANSCRIPT

Host: Mark Egner, Senior Manager, Accenture Security
Guest: Malek Ben Salem, Senior Principal, Cybersecurity R&D, Accenture

Mark: Hello. And welcome to On the Platform where we are talking to the most influential and innovative thinkers in platform technology on the hottest topics and trends. My name is Mark Egner, Accenture's Security Lead for our trust agenda in our software and platform client sector. Trust sits at that union of security, privacy, compliance, and fraud we see in the market today, and On the Platform we're going to explore bias and specifically the role it has in managing trust and safety of connected devices. I'm joined today by Malek Ben Salem, Senior Principal leading cybersecurity R&D out of Accenture's Washington, DC Cybersecurity Lab and a thought leader in this space. I've had the pleasure of working with and hearing Malek speak on these amazing topics so I've really happy to have her here with us today. Malek thanks so much for joining the program.

Malek: Thank you Mark.

Mark: So let's get started with a little context if you could please. Smart speakers seem to be everywhere and at holiday time most of us will give or receive maybe one of these. They understand us. They listen to us, but it can also be a little bit creepy. Can you just start with how these devices work, explain that a little bit?

Malek: Sure. Yeah, smart speakers are these voice-controlled devices that generally rely on accurate speech recognition for correct functionality. And as you mentioned they are constantly listening to us. They are recording everything we say and they upload whatever they record every three seconds to the Cloud.

Mark: Wow, I didn't know that.

Malek: And you know they are popular within the U.S. population. We know that 39-million Americans are using a smart speaker as of January 2018 and that's growing.

Mark: That's an incredible number. I had no idea. That many? That's more than one in ten.

Malek: Exactly. Yeah.

Mark: And so there's a lot of people using these. There's a lot of data flowing. It's highly connected. It makes sense. How is it that these devices actually understand how we are all communicating with them?

Malek: Yeah, so they have these speech recognition models that are machine learning models that get trained with samples of speech to recognize what people are saying. So they have an acoustic model and then a speech to text model to make that translation. And these models get trained with samples of regular speech from the average population which

means that ideally you want to have an unbiased dataset to train these devices so that you end up with an unbiased system that recognizes people speaking regardless of which gender they have or which demographic group they belong to.

Mark: So our clients that make these devices they make the device, they set-up the machine learning and they have to train these things with real humans' voices, yeah?

Malek: Mmhmm, exactly. Exactly. And it's important again to have an unbiased training data and in this context, in this statistical context when we are talking about biased we really mean - or unbiased – we mean that the sample that we are taking to train the model actually reflects the entire population.

Mark: Right, and that's hard to come by, right? So as you are making these devices there isn't just one of everybody so I would imagine they have to sort of chop it up and get it as good as possible, yeah?

Malek: Exactly. And it's easier to do for genders, right, training, these devices to recognize male voices versus female voices but it gets trickier if we are talking about different accents within the U.S., people from different regions because reflecting that distribution of the population not only is harder, we don't even know how people are distributed, but also if you train let's say with 50 percent of the data from people living in the southern region and 20 percent for people living in the east with an eastern accent that discrepancy in the samples that are used for training will result in a biased system. And a biased system in this case means that it does not perform in the same way at the same accuracy for people belonging to different demographics.

Mark: Interesting. So therein where there's an imperfect model I'm guessing we're getting to the point where any crack in the armor means it's susceptible to some type of cyber act of some sort, and I'm imagining you are able to tell us what's new about cyber attacks given so many people have these near-perfect or imperfect models in their hands right?

Malek: Exactly, yeah. So in this context bias is not just a problem with usability meaning that these systems do not perform as well for people having a certain accent. But beyond that it exposes that particular population to more cyber-attacks, and the way this happens is through what is called skill-squatting.

Mark: Skill-squatting. I haven't heard that term. Skill-squatting, interesting.

Malek: That is a pretty new term. So if you think about Alexa, which is a speech recognition engine that works typically with Amazon, the Amazon Echo family, where skills are kind of apps that get trained and that users interact with through voice commands. What happens is that these skills obviously have names, right, but if the name of the skill gets mispronounced or if Alexa misinterprets the name of this skill Alexa recognizes or works by recognizing this skill name that the user wants to interact with. If a person has a certain accent and pronounces that skill name in a different way then Alexa may confuse it with another skill that's available on the store and may misdirect that user to the wrong skill.

Mark: So that's the imperfection in the model, producing a result we didn't want, but is that necessarily dangerous?

Malek: Yeah. So adversaries can in advance identify which errors Alexa may make based on knowing how people from a certain region speak. So these errors can be predictable and there have been studies that were performed that showed that these errors are predictable. Alexa makes the same errors interpreting or misinterpreting peoples' words if they speak with a certain accent over and over. So if the adversary knows in advance that this skill name will be misinterpreted in this way they can create another skill that sounds similar to the way it gets pronounced and they can upload it to the skill store, to the app store and then rely on the fact that Alexa will misinterpret the voice command when it gets issued by those people from a certain region speaking with a certain accent. All of those people or any significant portion of them will get misdirected to this other

malicious skill and the adversary can harness now the interaction with that skill to fool the users to share their credentials. For instance, if we are talking about skill that requires those types of credentials.

Mark: Amazing. So it truly is a magic word that causes this intentional redirect and because the magic word or phrase is predictable, bad actors can target and they can get after these things, so that's skill-squatting. That's amazing. I never would have put those pieces together. So when we think about that as an exposure what are we talking to our clients about? If we are reusing these devices and machine learning and speech recognition inside of business processes, the insights that you've just mentioned here, I mean what do we talk to the clients about to talk with them and how it's used or what we can do about it?

Malek: There are a number of things that can be done. You know the obvious one is for Amazon to make some vetting as these skills get contributed and uploaded to make sure that the way they get pronounced is not - that there are no skills that have similar names.

Mark: Right, so a naming convention control. It makes sense.

Malek: Exactly. As it is there is one skill that sounds like cat facts, right? There is almost 400 skills like that, 400 pairs of skills with different names that sound similar that exist today for Alexa. So certain checks can be done in order to eliminate the room for skill-squatting. There's also work that can be done in terms of the architecture and the flow of how people interact with these skills. So as of April 2017 Alexa used to require, or Amazon used to require users to enable these skills to their account in a manner similar to downloading a mobile app onto your personal device.

Mark: To authorize it, to use and direct and redirect, yeah. Got it.

Malek: Exactly. However, since then they have done away with that process so now anybody can interact with any skill on the Cloud. So that prior authorization doesn't happen which creates more room for misdirecting the user to the wrong skill.

Mark: Of course. That makes sense. So I don't have to off-in to some central control. I can just go one point to the next point. That's amazing. Yeah. Anything else?

Malek: Yeah, there are other things that can be done in terms of moving this speech recognition process or these speech recognition models to the end point device so that the model gets trained and created with the final user sample, right. It gets personalized to their own voice, to their own accent, and all the speech recognition and the translation from speech to text would happen on the end point device. That's obviously better in terms of privacy but it's also better in terms of security.

Mark: We're just not there yet in terms of local computing power and the way these models work, is that right?

Malek: That is right, yeah. So on the Cloud we are dealing with huge deep learning networks, but there's some research going on to compress these networks so that they can be run on an end point basis.

Mark: So this is great, Malek. What you've done from Cyber Labs is look at something that's in everyone's hands today that we can all relate to, looked at it deeply to understand how it works and identified vulnerabilities that this capability has that we might not all think of. And you've given us three really great things to talk to clients about, that these smart connected devices we should be thinking in terms of the controls around naming convention of how they recognize skills and things you want them to do. You've talked a little bit about the architecture of how those skills relate to authentication source to connect to each other to reduce this skill-squatting. And you've also mentioned about product conversations when product builders are enacting these kinds of skills, ways to be thinking about the trend ahead where the power moves to the device and unique things that could be done.

Unfortunately that is all the time we have today. We hope you've enjoyed this episode. Please help us get the word out and be sure to subscribe, share, rate, and review our series. Again, Malek I want to thank you personally. It's always fun to have you. Thanks for joining. We would love to hear from all of you and hope you tune-in again for the next episode of *On the Platform.*

[End of recording.]

Copyright © 2019 Accenture
All rights reserved.

Accenture, its logo, and High Performance Delivered are trademarks of Accenture.