Are you listening? – an observational wake word privacy study

Marcia Combs and Casey Hazelwood
Cybersecurity Management Program, Murray State University, Murray, Kentucky, USA, and
Randall Joyce
Cybersecurity and Network Management Program, Murray State University, Murray, Kentucky, USA

Received 5 December 2021
Revised 18 May 2022
Accepted 19 May 2022

Abstract

Purpose – Digital voice assistants use wake word engines (WWEs) to monitor surrounding audio for detection of the voice assistant's name. There are two failed conditions for a WWE, false negative and false positive. Wake word false positives threaten a loss of personal privacy because, upon activation, the digital assistant records audio to the voice cloud service for processing.

Design/methodology/approach – This observational study attempted to identify which Amazon Alexa wake word and Amazon Echo smart speaker resulted in the fewest number of human voice false positives. During an eight-week period, false-positive data were collected from four different Amazon Echo smart speakers located in a small apartment with three female roommates.

Findings – Results from this study suggest the number of human voice false positives are related to wake word selection and Amazon Echo hardware. Results from this observational study determined that the wake word Alexa resulted in the fewest number of false positives.

Originality/value – This study suggests Amazon Alexa users can better protect their privacy by selecting Alexa as their wake word and selecting smart speakers with the highest number of microphones in the far-field array with 360-degree geometry.

Keywords Wake word engine, Smart speakers, Amazon Alexa, Personal privacy

Paper type Research paper

1. Introduction

Digital voice assistants are becoming a common trend for home automation, and there are several vendors such as Amazon, Google, Apple and Xiaomi. These different vendors' voice assistants use wake word engine (WWE) algorithms to constantly monitor surrounding audio for detection of the voice assistant's name. Amazon is one of the more popular brands on the market, and in 2019, Amazon reported that there were close to 100 million devices sold that utilize the Alexa voice assistant (Bohn, 2019). Once the assistant's name, Alexa," "Echo," "Hey Siri," also known as the wake word, has been identified, the digital voice assistant is activated and streams the subsequent verbal audio (commands) to a cloud service. The cloud service then uses speech recognition and natural language understanding algorithms to translate and execute the voice command, such as turning on the lights. WWE algorithms rely upon far-field microphone arrays and digital signal processing (DSP) to collect, condition

Organizational Cybersecurity Journal: Practice, Process and People
Vol. 2 No. 2, 2022
pp. 113-123
Emerald Publishing Limited
e-ISSN: 2635-0289
p-ISSN: 2635-0270
DOI 10.1108/OCJ-12-2021-0036

and digitize the analog voice signal. The voice signal is then compared to a stored wake word pattern or a profile from a trained neural network. If the digitized voice pattern matches with high confidence to the algorithm's stored wake word profile, then the digital assistant is activated and carries out the user's commands. WWEs are typically stored and executed locally on the smart device hardware (Kinsella, 2018). However, wake word verification can be performed locally and verified again in the cloud, such as in the case of Amazon Alexa wake word cloud verification (Karczewski, 2017).

The WWE's efficacy depends on the hardware's ability to capture the human voice. Digital assistants embedded in devices such as Amazon's Echo smart speakers rely upon far-field microphone arrays to capture the human voice. The array's ability to capture a quality input signal (human voice) is influenced by the acoustic environment, microphone design and DSP (Lu, 2017). Home acoustic environment variable examples are the user's proximity to microphone array, level of ambient noise, signal (voice) reverberation and the output signal of the smart speaker itself (Haeb-Umbach *et al.*, 2019). WWE designers cannot control the user's proximity to the microphones, the level of noise or reverberation within the home environment. However, they can integrate hardware techniques such as beamforming, increased number of microphones, array geometry design and the use of the DSP technique acoustic echo cancellation to reduce noise within the home acoustic environment (Connor, 2018), thereby improving the accuracy of the WWE.

WWE efficacy is also dependent on the proper selection of wake words. What makes a good wake word? How do WWE developers choose a wake word that improves the engine's accuracy? According to voice and artificial intelligence platform developers, wake words should be unique, consist of at least six phonemes, contain words with diverse sounds, easily pronounced and contain a gender-neutral name. It is recommended to avoid using brand names or phrases that rhyme to reduce the number of false wakes. Environmental noise such as brand names used in TV commercials may trigger the wake engine. In addition, introduction words such as "hi" or "hey" can be added to short wake words to reduce the number of false wakes (Scates, 2019 Picovoice, 2020). Once the chosen wake word is spoken and captured by the far-field microphone array, conditioned and digitized by DSP, it is then compared to a wake word pattern based on a template or profile. Using artificial intelligence learning techniques, automatic speech recognition frameworks create an acoustic profile that matches the wake word input signal or voice pattern (Ge and Yan, 2017). Over time, as speech patterns are learned, the acoustic profiles more closely match the digital assistant user's voice pattern, thereby improving the accuracy of the WWE.

2. Security concerns with wake words

WWEs and microphone arrays can fail to detect the wake word. There are two failed conditions for a WWE, false negative and false positive. A false negative is when the wake word is spoken, but the WWE fails to recognize the wake word, and the digital assistant is not activated. A false positive is when the WWEE inaccurately detects the wake word and activates the digital assistant. One of the first explorative works on the confusion of wake words was done at Georgetown University, where they looked at Google voice assistance on a smartphone. What was found was that the phrase "Cocaine noodles" would be misinterpreted as "Ok Google" (Vaidya *et al.*, 2015). This would allow the user to exploit this behavior to execute unauthorized commands (Vaidya *et al.*, 2015). For privacy advocates, false positives are more concerning because, upon activation, the digital assistant streams and records subsequent audio to the voice cloud service for processing. Unless users are provided visual notification of streaming, such as Amazon's Echo's blue spinning light ring, they are entirely unaware their voices have been recorded. These inadvertent recordings contain splices of private conversations among family members, including children talking, adult private

conversation or interactions between family members; these inadvertent recordings offer breadcrumb views into the private lives of digital assistant consumers. Google and Amazon have been criticized for their lack of privacy controls for protecting these user audio recordings. In 2018, an Amazon Echo recorded and sent a couple's conversation to a member of their contact list without their knowledge (Wamsley, 2018). In 2019, a language reviewer hired by Google leaked 1,000 digital assistant recordings to the Belgian broadcaster VRT who contacted the Google Home user and notified them of the privacy leak (BBC News, 2019). Researchers have also found an attack called skill squatting that leverages transcriptions errors of a list of similar-sounding words to current Alexa skills, and then when the similar-sounding words are executed, it routes to the malicious skill with a similar name (Kumar *et al.*, 2018). There have also been several other voice assistants that exploit how the assistant invokes the skill (Zhang *et al.*, 2019; Zhang *et al.*, 2018). Despite privacy concerns, Amazon will continue to collect consumer recordings "by keeping this data. . .it improves the service materially" (Guthrie, 2019) by providing neural network training data sets.

As mentioned earlier, false positives have been identified as a threat to digital voice assistant users' privacy. Consumers have no control in the development of WWE algorithms, far-field microphone design or acoustic profiles. However, in a select number of digital assistant products, users can control the selection of the wake word. With an estimated global deployment of 8 billion digital voice assistant-enabled devices by 2023 (Voicebot.ai, 2019), the question is – which wake word results in a fewer number of false positives? If users select the wake word that results in fewer false positives, they potentially limit their exposure to users' personal identifiable information. Therefore, the purpose of this observational study was to identify which wake word resulted in the fewest number of human voice-initiated false positives (false wake) by collecting and analyzing the number of false positives from four smart speakers embedded with personal digital assistants.

3. Security vulnerabilities

Smart speakers have also been referred to as home digital voice assistants (HDVAs) (e.g. Amazon, Alexa, Google Home) and have been very popular over the years, and most are controlled through voice commands utilizing wake words. Many manufacturers continue to work to advance HDVA's hardware and third-party voice developers. These third-party applications vary from playing music to online shopping to physical controls in the home. However, with the HDVAs having these capabilities, it creates security and privacy threats because of the openness of these applications. What has been found with the HDVAs and wake words is that they suffer from three significant security vulnerabilities.

The first security vulnerability that Alexa and other HDVAs are victims of is weak single-factor authentication. The single-factor authentication is based on a wake word to authenticate a user that will execute the command after speaking the wake word. Alexa's voice service does not support voice authentication, so any user can issue the command if the correct wake word is used (Lei *et al.*, 2017). Alexa's wake words are publicly known, and that would allow a malicious user to authenticate and issue a command with ease (Lei *et al.*, 2017). The weak single-factor authentication that Alexa and the other HDVAs have is a critical vulnerability for the vendors to resolve. However, one HDVA vendor has started working on single-factor authentication. Google has developed a new voice service called Voice Match. Voice Match allows users to train Google Assistant to recognize your voice to authenticate commands (Charlton, 2020). The Voice Match service allows for up to six users' voices to be added to the service. The Voice Match services are trained by recording the user's voice locally and sending it to a Google server to validate the voice against a voice model created during the user's voice enrollment process. Once the voice is validated or rejected, the recording is immediately deleted from the Google server. Google acknowledges that the service can still be tricked into a recording of the user's voice if it can match

the voice model for that user (Charlton, 2020). Implementing a voice verification service into the HDVA is a step in the right direction to strengthen the single-factor authentication, but HDVAs will have to add more authentication and validation steps in the future.

The second security vulnerability is that the HDVAs and wake words suffer no physical presence-based access control. If a person or device can produce a sound pressure level of 60 dB, it can command the HDVAs using the wake words (Lei *et al.*, 2017). This means that an adversary does not need physical access to the HDVA to send a wake word command. They could be in a different area and send the wake word to the HDVA. Having no physical presence-based access control also means adversaries could send commands through electronic devices to command the HDVA (Lei *et al.*, 2017). With both methods, an adversary could send commands to third-party applications configured with the HDVAs for information and personal gain. Since most third-party applications have known commands, adversary users could use them to order or unlock doors through either technique. Implementing a physical presence-based access control for the HDVAs will be critical in the future to ensure privacy for the users.

The final vulnerability that HDVAs suffer from is insecure third-party application and skill access. With a plethora of information about HDVAs and all the third-party applications and smart devices associated with them, it enables the adversary to have a decent chance for success in sending the correct wake word and command for the smart device. For example, an adversary could issue a command like "Alexa unlock the front door" if Alexa has been connected to the smart lock system. Alexa allows the user to change the labels for these smart devices, but it is not mandatory and will allow the user to keep the default label usually used in that device's documentation. Additionally, companies like Schlage have added authentication measures into smart lock systems where the user must configure a pin code that has to be used for the door to be unlocked after the Alexa command is issued (Wolpin, 2021). However, not all companies have implemented measures like these with smart devices with cyber-physical connections.

Other exploits that have been used with Alexa's third applications and skills are Voice Squatting and Voice Masquerading. Voice Squatting is where the adversary sets the invocation of a malicious skill to something like an actual skill (Kumar *et al.*, 2018; Corbett and Karafili, 2021). Often, this is done by removing words or using homonyms, and then when the user starts communicating with the service, the user believes it is legit (Kumar *et al.*, 2018; Corbett and Karafili, 2021). Voice Masquerading works similarly, but it focuses on when a user tries to switch between skills (Mitev *et al.*, 2019; Corbett and Karafili, 2021). If a user does not preface their switch command with the wake word, then they will continue to use the current skill, and that is where the exploit can occur, and the malicious skill will emulate what the real skill does in the hope of collecting user data (Mitev *et al.*, 2019; Corbett and Karafili, 2021). Another exploit that is like Voice Masquerading is Voice Surfing. Voice Surfing is an exploit that allows the adversary to interact with the HDVAs from a long distance (Yan *et al.*, 2020; Corbett and Karafili, 2021). It has been proven that malicious skills could also be built for just the purpose of scrapping information from users. For example, in a research study conducted at the University of Southampton, they created a skill called "County Facts" that was used for payment detail harvesting that would request the user to add payment information (Corbett and Karafili, 2021). Knowing how third-party applications and skills can be exploited is critical in learning and knowing how to defend HDVAs and will promote further research in the future. Understanding these security vulnerabilities with HDVAs and wake words is critical in ensuring a user's privacy and security. That is why this study is investigating what false positives associated with wake words cause the smart speakers to listen in and record users without them being aware of it.

4. Methodology

During an eight-week period, January 19, 2020, through March 14, 2020, four Amazon Echo smart speakers, embedded with the digital personal assistant, Alexa, were installed in an

open-concept apartment. The apartment residents (participants) and guests were instructed not to purposefully interact with the digital assistant but carry out their lives and disregard the presence of the smart speakers. Participants (three women) engaged in normal everyday life activities, and they went to work, school, cooked and entertained friends. Since the participants did not purposefully interact with the smart speakers, all-digital assistant voice commands logged by the smart speakers were categorized as false positives. No participant personal information was collected during this study.

As shown in Table 1, four different smart speaker models were placed in an 800 square foot, open-concept apartment. As shown in Figure 1, highlighted in red, the Amazon Echo smart speakers were located near the form door within the living room. The smart speaker network topology was a typical home smart speaker installation (see Figure 2) using default smart speaker configuration.

During the eight-week data collection period, all voice activity from each smart speaker was transcribed, logged (time and date) and categorized using the Amazon Alexa application. Since the participants did not purposefully interact with the smart speakers, all activities logged by the smart speakers were categorized as false positives. Voice activity was logged and categorized as the original recording, duplicate recording, TV audio, human audio and continuation of an earlier audio recording. Within the Amazon Alexa application, some logged voice activity exhibited anomalies. Some Amazon Echo voice activity was logged twice within the Amazon Alexa application for unknown reasons. For example, a false positive voice recording such as, "and then he said," was listed twice in the Alexa application – two instances of the same voice activity. The first voice activity log was categorized as an original recording, and the second voice activity was categorized as a duplicate. Another experienced anomaly was that some voice logs were displayed twice, with the first audio log containing the beginning of the voice recording and the second log containing the continuation of the voice recording. For example, using the earlier voice command example, the first log contained "and then" and the second log contained "he said." The second or continuation voice activity was categorized as a continuation. In addition, some Amazon Echo logged voice activity did not contain human or TV voice recording, but sounds within the environment such as chopping vegetables were categorized as "other." The purpose of this observational study was to identify which wake word resulted in the fewest number of human voice-initiated false positives (false wake); therefore, only human voice, original false positives, were analyzed.

In an attempt to account for the smart speakers' hardware influence on false positives, at the beginning of each week, the Amazon wake words Alexa, Echo, Computer and Echo were rotated between the four different Echo smart speakers. Thus, each Amazon smart speaker was configured with all four different wake words for a total period of two weeks. In addition, each week, Amazon Echo smart speaker firmware versions were documented to determine firmware impact on false positives.

5. Findings

As shown in Table 2, a total of 225 original false positives were recorded between January 19, 2020, through March 14, 2020 (eight weeks), with 64% of the total recordings categorized as a human voice. Since this study focused on user personal privacy, false positives recordings

Smart speaker name	Smart speaker device	Model number
Echo Show 5	Echo Show 5	H23K37
Echo Dot 2	Echo Dot 2nd generation	RSO3QR
Echo Dot 3	Echo Dot 3rd generation	C78MP8
Echo 3	Echo 3rd generation	R9P2A5

Table 1. Smart speaker equipment list

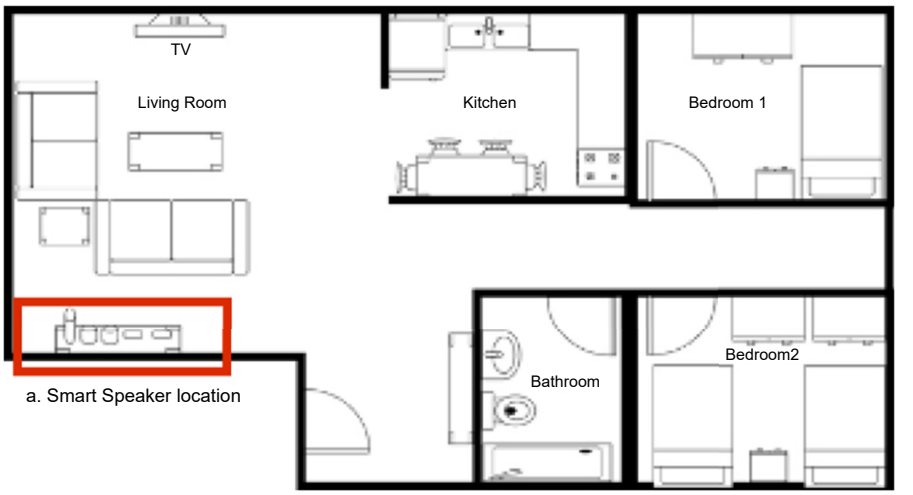

Figure 1.
Participants' apartment layout

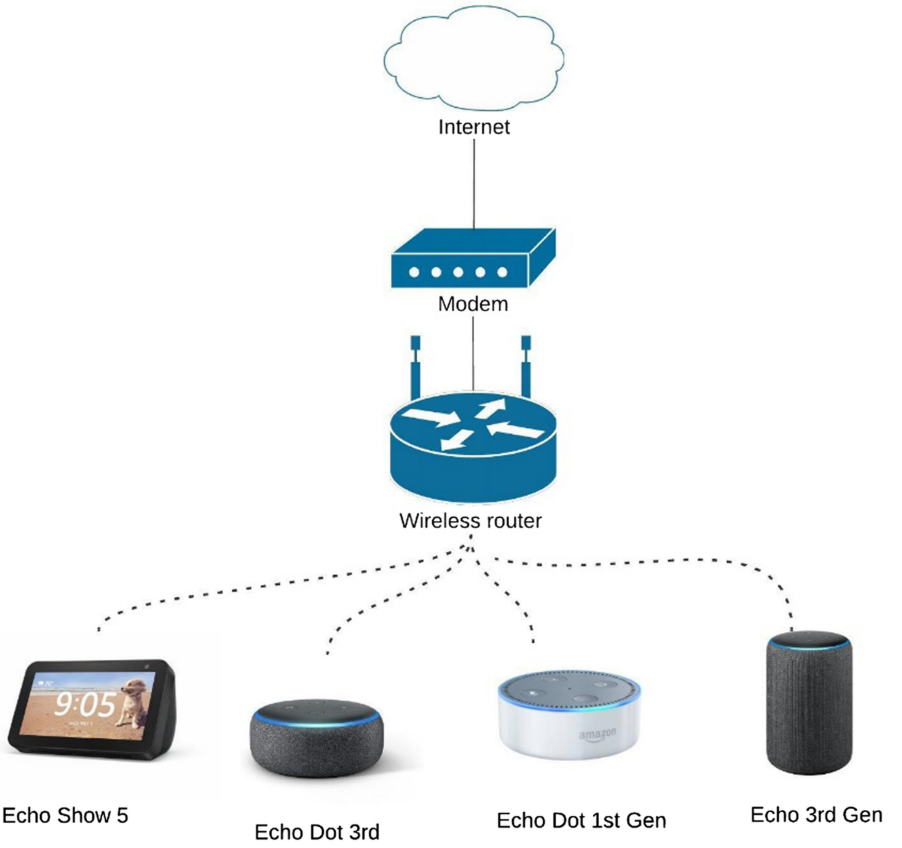

Figure 2.
Smart speaker network topology

categorized as repeats, TV recordings or other household sounds were omitted from data analysis, leaving 144 original human voice false positive observations for analysis.

Which Amazon wake word resulted in the fewest number of original, human voice false positives? As shown in Table 3, the wake word *Alexa* resulted in the fewest number of wake word false positives with a value of 6, while the wake word *Amazon* exhibited largest number of false positives with a value of 65. Which Amazon smart speaker experienced the fewest number of original, human voice false positives? As highlighted in Table 3, the Amazon Echo Dot 2 (2nd generation) smart speaker resulted in the fewest number of human voice false positives, with a count of 20, while the Echo Show 5 tallied the largest number of false positives with a count of 56.

The data in Table 3 support WWE developers' recommendation of choosing a wake word with at least six phonemes, uniqueness and avoidance of brand names. The wake word *Alexa* with 6 phonemes exhibited the fewest false positives. The wake words *Amazon* and *Computer* with phonemes of 6 and 8, respectively, using the recommended number of phonemes, were common and brand-specific names. These wake words, which are found in everyday conversations, resulted in the highest number of false positives. While the wake word *Echo* with 3 phonemes followed wake word recommendations of uniqueness and short recorded 30 false positives. These results are possible due to the wake word *Echo* sharing the phoneme *k* with one of the participants' names.

These findings suggest that the microphone array design influences wake word accuracy and automatic speech recognition. The Echo Dot 2 (2nd generation Echo Dot) and Echo 3 (3rd generation Echo), both with a seven far-field microphone array with 360-degree geometry, experienced the fewest false positives, with 20 and 31 false positives, respectively. The Echo Dot 3 (3rd generation Echo Dot) with a four-microphone far-field array with 360-degree geometry recorded 37 false positives. With a twin far-field microphone array, the Echo Show 5 recorded the highest number of false positives, with a frequency of 56. However, since the exact micro-electromechanical system (MEMS) microphone or DSP chip(s) used in the Amazon Echo are unknown, these findings are inconclusive. Additional research is required to verify the effect of microphone array design on wake word accuracy.

During the collection period, by rotating the Amazon wake words Alexa, Amazon, Computer, and Echo among the four different Echo smart speakers provided false positive

Type of audio type	Count
Original	195
Human voice	144
TV	46
Other	5
Repeat	30
Total	225

Table 2. Count of wake word false positives by audio type

Device	Wake word				
	Alexa	Amazon	Computer	Echo	Total
Echo Dot 2	2	9	2	7	20
Echo 3	1	7	13	10	31
Echo Dot 3	2	8	21	6	37
Echo Show 5	1	41	7	7	56
Total	6	65	43	30	144

Table 3. Count of original, human voice wake word false positives by device and wake word

observations across a diversity of hardware and firmware configurations. Each smart speaker was configured with one of the four wake words for a period of two weeks. Meaning, weeks 1–4 wake word assignments were the same as weeks 5–8. In addition, the smart speakers' firmware updated automatically during the collection period.

How did the count of original, human voice false positives change over time? As shown in Tables 4 and 5, the total count of false positives decreased by 60% from weeks 1–4 to weeks 5–8, from 103 false positives to 41 false positives, respectively. As shown in Tables 4 and 5, weekly original human voice false positives by device and wake word also decreased. The data show the number of false positives decreases over time regardless of chosen wake word or smart speaker hardware ,suggesting the digital assistant's artificial intelligence and WWE learned the participants' voice profile. However, since the details of each smart speakers' artificial intelligence or neural network algorithms are unknown, these findings are inconclusive, and additional research is required to determine if wake word false positives decrease over time.

How did the original human false positive count change by the device over the 8-week collection period? During week 3, Echo Dot 3 configured with the wake word *Computer* and Echo Show 5 configured with the wake word *Amazon* experienced the highest number of false positives, with values of 15 and 35, respectively (see Figure 3). However, in week 7 (same wake word configuration as week 3), both Echo Dot 3 and Echo Show 5 recorded fewer false positives, with a count of six for each device (see Figure 3).

What happened during week 3 that caused the increase in original, human voice false positives? Echo Show 5 underwent a firmware upgrade from version 3322491268 to version 3322491269 during week 3, suggesting the firmware upgrade resulted in an unusually high number of false positives. However, in week 6, Echo Show 5 underwent another firmware version upgrade, but without a high number of false positives. Therefore, it is unclear if Echo Show week 3 firmware upgrade affected the number of false positives or if there was more human activity in the home during this time, as reported by the participants. Additional research is required to verify the effect of firmware upgrades on wake word accuracy.

The results of this observational study suggest three main findings: (1) the number of false positives is related to wake word, (2) number of false positives is related to Amazon Echo hardware and (3) false positives decrease over time.

Table 4.
Count of original, human voice wake word false positives by device and wake word for weeks 1–4, January 19–February 15, 2020

Device	Wake word				
	Alexa	Amazon	Computer	Echo	Total
Echo Dot 2	2	6	–	4	12
Echo 3	1	2	10	8	21
Echo Dot 3	–	3	15	6	24
Echo Show 5	1	35	4	6	46
Total	4	46	29	24	103

Table 5.
Count of original, human voice wake word false positives by device and wake word for weeks 5–8, February 16–March 14, 2020

Device	Wake word				
	Alexa	Amazon	Computer	Echo	Total
Echo Dot 2	–	3	2	3	8
Echo 3	–	5	3	2	10
Echo Dot 3	–	6	3	1	10
Echo Show 5	2	5	6	–	13
Total	2	19	14	6	41

6. Discussion

Consumers have not only adopted smart speakers and smart home technology, but they have embraced them with passion, for it is estimated that 333 million homes contain some type of smart home product, and 186 million smart speakers were shipped worldwide in 2021 (Laricchia, 2022). By their design, millions of consumers have unknowingly installed smart speaker eavesdropping devices in their homes. These devices inadvertently record portions of private conversations and transmit them to the cloud for processing and storage. Who is responsible for the security and privacy of these snippets of private conversation? The consumer? The smart speaker manufacturers? The results of this study demonstrate the security and privacy of smart speaker voice recordings is the responsibility of the consumer, the smart speaker design engineers and WWE software developers.

Consumers can protect their privacy by reducing the number of inadvertent voice recordings (false positives) made by smart speaker. Three key findings from this study demonstrate how consumers can potentially reduce the number of inadvertent recordings captured by Amazon Echo smart speakers. This study demonstrated (1) the use of the wake word *Alexa* resulted in the fewest false positives regardless of smart speaker hardware configuration, (2) smart speakers with the highest number of microphones in the far-field array arranged in a 360-degree geometry recorded the fewest number of false positives and (3) false positives decreased over time, suggesting smart speaker artificial intelligence was learning voice patterns. Therefore, consumers can protect their privacy by selecting the *Alexa* wake word, purchasing smart speakers with the highest number of microphones in a 360-degree geometry pattern and opting for voice recognition profiles.

WWE software developers contribute to the security and privacy of consumers' voice recordings by selecting engines wake word based on best practices of using unique words, use of at least six phonemes and avoid common words and brand names. Smart speaker design engineers contribute to consumer security and privacy by designing products with six or more microphones arranged in a 360-degree geometry pattern.

7. Conclusion

While the three key findings from this study demonstrated security and privacy efforts for consumers regarding smart speakers, the findings are incomplete due to the limitations of the study. Future smart speaker security and privacy research studies could include a larger number of more diverse participants. Future studies could include participants with different dialects, accents, varying tones and gender diversity. Another limitation of this study was the

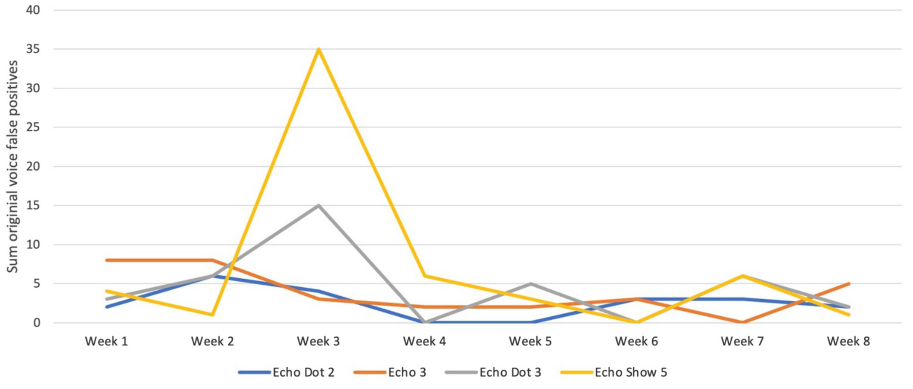

Figure 3.
Weekly count of original human voice false positive by the device

inability of the researchers to control the Echo's firmware versions. Not all the Echo models used the same firmware version; for example, Echo Show used a different firmware version than the Echo Dots. In addition, as shown in week 3, Echo Show 5 underwent a firmware version, but the other Echo hardware platforms did not update their firmware. Since Amazon does not release a firmware version fixes and features list, the effect of different firmware versions is unknown in this study. Another limitation of this study is the modifications of Amazon artificial intelligence voice services during this data collection period.

Amazon continues to develop smart speaker listening features such as Home Guard, a free service that constantly listens for smoke alarms or glass breaking. Amazon's 2019 U S. Patent application 20190156818 adds features to Amazon Alexa WWEs to begin recording audio before users speak the wake word, allowing the voice assistant to review longer voice tracks of previously recorded audio looking for its name (Dellinger, 2019). For these reasons, it is essential to continue research in security and privacy concerning smart speaker hardware, firmware, WWEs and wake word selection.

References

BBC News (2019), "Google probes leak of smart speaker recordings", available at: https://www.bbc.com/news/technology-48963235.

Bohn, D. (2019), "Amazon says 100 million alexa devices have been sold", Jan. 2019, available at: https://www.theverge.com/2019/1/4/18168565/ (accessed 31 July 2020).

Charlton, A. (2020), "Google Assistant and Voice Match: AI voice recognition features and security explained", available at: https://www.gearbrain.com/google-assistant-voice-match-explained-2647642657.html (accessed 5 December 2021).

Connor, Kevin (2018), "Smart speakers: audio design rules for voice-enabled devices", *AudioXpress*, available at: https://audioxpress.com/article/smart-speakers-audio-design-rules-for-voice-enabled-devices (accessed 31 May 2020).

Corbett, J. and Karafili, E. (2021), "Private data harvesting on Alexa using third-party skills", *International Workshop on Emerging Technologies for Authorization and Authentication*, Springer, Cham, pp. 127-142.

Dellinger, A.J. (2019), "Amazon considered letting Alexa listen to you without a wake word", available at: https://www.engadget.com/2019-05-23-amazon-alexa-recording-before-wake-word-patent.html?guccounter=1 (accessed 1 June 2020).

Ge, F. and Yan, Y. (2017), "Deep neural network based wake-up-word speech recognition with two-stage detection", *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, pp. 2761-2765, doi: 10.1109/ICASSP.2017.7952659.

Guthrie, G. (2019), "Amazon says it won't stop recording user interactions with Alexa", available at: https://www.consumeraffairs.com/news/amazon-says-it-wont-stop-recording-user-interactions-with-alexa-093019.html.

Haeb-Umbach, R., Watanabe, S., Nakatani, T., Bacchiani, M., Hoffmeister, B., Seltzer, M.L. and Souden, M. (2019), "Speech processing for digital home assistants: combining signal processing with deep-learning techniques", *IEEE Signal Processing Magazine*, Vol. 36 No. 6, pp. 111-124.

Karczewski, T. (2017), "Cloud-based wake word verification improves 'alexa' wake word accuracy on your AVS products", available at: https://developer.amazon.com/blogs/alexa/post/b136b3e7-0ba8-4589-aaf9-2a037fc4e9c9/cloud-based-wake-word-verification-improves-alexa-wake-word-accuracy-on-your-avs-products.

Kinsella, B. (2018), "Voicebot podcast episode 24", Todd Mozer CEO Sensory Discusses Neural Nets, Wake Words and Two Decades of Voice Technology, available at: https://voicebot.ai/2018/01/09/voicebot-podcast-episode-24-todd-mozer-ceo-sensory-discusses-neural-nets-wake-words-two-decades-voice-technology/.

Kumar, D., Paccagnella, R., Murley, P., Hennenfent, E., Mason, J., Bates, A. and Bailey, M. (2018), "Skill squatting attacks on amazon alexa", *27th {USENIX} Security Symposium ({USENIX} Security 18)*, pp. 33-47.

Laricchia, F. (2022), "Topic: smart speakers", *Statista*, available at: https://www.statista.com/topics/4748/smart-speakers/#dossierKeyfigures (accessed 13 May 2022).

Lei, X., Tu, G.H., Liu, A.X., Ali, K., Li, C.Y. and Xie, T. (2017). "The insecurity of home digital voice assistants–amazon alexa as a case study", arXiv preprint arXiv:1712.03327.

Lu, J. (2017), "Can you hear me now? Far-field voice", Medium, available at: https://towardsdatascience.com/can-you-hear-me-now-far-field-voice-475298ae1fd3 (Retrieved 31 May 2020).

Mitev, R., Miettinen, M. and Sadeghi, A.R. (2019), "Alexa lied to me: skill-based man-in-the-middle attacks on virtual assistants", *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security*, pp. 465-478.

Picovoice (2020), "Picovoice.ai", available at: https://picovoice.ai/blog/tips-for-choosing-a-wake-word/ (accessed 31 May 2020).

Scates, Karen (2019), "The art and science of creating a custom wake word for your brand - speech-to-meaning blog", *Speech-to-Meaning Blog*, available at: https://voices.soundhound.com/the-art-and-science-of-creating-a-custom-wake-word-for-your-brand/ (accessed 31 May 2020).

Vaidya, T., Zhang, Y., Sherr, M. and Shields, C. (2015), "Cocaine noodles: exploiting the gap between human and machine speech recognition", *9th {USENIX} Workshop on Offensive Technologies*, ({WOOT} 15).

Voicebot.ai. (2019), "Number of digital voice assistants in use worldwide from 2019 to 2023 (in billions) * [Graph]", *Statista*, available at: https://www-statista-com.ezproxy.waterfield.murraystate.edu/statistics/973815/worldwide-digital-voice-assistant-in-use/ (accessed 28 February 2020).

Wamsley, L. (2018), "Amazon echo recorded and sent couple's conversation — all without their knowledge", available at: https://www.npr.org/sections/thetwo-way/2018/05/25/614470096/amazon-echo-recorded-and-sent-couples-conversation-all-without-their-knowledge.

Wolpin, S. (2021), "Review: schlage encode smart wi-fi deadbolt door lock", *Gearbrain*, available at: https://www.gearbrain.com/review-encode-smart-lock-schlage-2638646827.html (accessed 5 December 2021).

Yan, Q., Liu, K., Zhou, Q., Guo, H. and Zhang, N. (2020), "Surfingattack: interactive hidden attack on voice assistants using ultrasonic guided waves", *Network and Distributed Systems Security (NDSS) Symposium*.

Zhang, N., Mi, X., Feng, X., Wang, X., Tian, Y.1 and Qian, F. (2018), "Understanding and mitigating the security risks of voice-controlled third-party skills on amazon alexa and google home", arXiv preprint arXiv:1805.01525.

Zhang, N., Mi, X., Feng, X., Wang, X., Tian, Y. and Qian, F. (2019), "Dangerous skills: understanding and mitigating security risks of VoiceControlled third-party functions on virtual personal assistant systems", *IEEE Symposium on Security and Privacy, Ser. SP '19*, IEEE, San Francisco, CA, pp. 1381-1396.

Corresponding author
Randall Joyce can be contacted at: rjoyce@murraystate.edu