Home Digital Voice Assistance: A Case Study

Atul kumar[1], Ankush Negi[2], Pallavi Gupta[3], Harendra Sharma[4]

Computer Science and Engineering Department
Dev Bhoomi Institute of Technology, Dehradun
(E-mail: atulmishra845@gmail.com [1] , ankushnegi5555@gmail.com [2], pallavigupta167@gmail.com [3], harendra.pth@gmail.com [4])

Abstract—Digital Voice Assistance like Alexa AI echo dot, Google mini etc. are getting popular today as they let you use compatible devices without taking out your phone and start an application. While an AI-powered machine makes life much easier, they have some serious issues on our privacy and security.

This paper discusses all the disadvantages of the voice-control system, Alexa and problems that can occur if it is handed in wrong hands. This paper also suggests some solutions by which we can protect the device from being misused.

Keywords—*Alexa, AI, Artificial Intelligence, HDVA's, IVA's, Google-Mini, Siri, Cortana.*

I. INTRODUCTION

Voice-based AI devices also known as Home Digital Voice Assistant (HDAV) are getting famous because of its amazing Voice User Interface (VUI) technique. They are mainly helpful for those people who cannot easily type or write on devices like mobile phones, tablets etc.

Digital voice assistance, a part of AI (Artificial Intelligence) is being part of our daily lives rapidly. This is the coming future of our generation. Artificial intelligence focuses on research and optimization, processing our natural languages & representation of our knowledge, making decisions based onour daily routine. AI stands different for a different public; it has a various number of definitions.

In today's era, AI is playing a vital role in our day to day life, solving issues like security and user friendliness will be the central focus on how to develop and integrate upcoming AI system.

Millions of Alexa devices are sold since the market of voice assistance devices came into existence, the main payoff is the opportunity to control mainly three important areas of markets i.e. automation in the home, automation in entertainment and the most important of one's daily life- Shopping.

Alexa word is derived from a Greek word ALEXANDRA-defender of men and it was named from his library which stores the knowledge of the ancient world. Alexa works for the sound whose sound pressure level is higher or more than 60DB. Alexa has inbuilt more than 500 skills (Skills refers to software programs in the words of Amazon).

Once a user speaks its key or can say password word i.e. Alexa, it gets started in listening mode. For convenience, user can change this key from the Alexa mobile App or by Alexa website. Alexa takes voice commands from its user and transfers that to voice processing cloud through the internet where our commands are simplified and specific tasks are performed.

Besides all the advantages and features provided by AI, there is always a concern of the dark side of these technologies, including possibility of decreasing jobs in society, privacy, security and the most important "THE SINGULARITY"- the point at which machine starts learning itself, without any human knowledge.

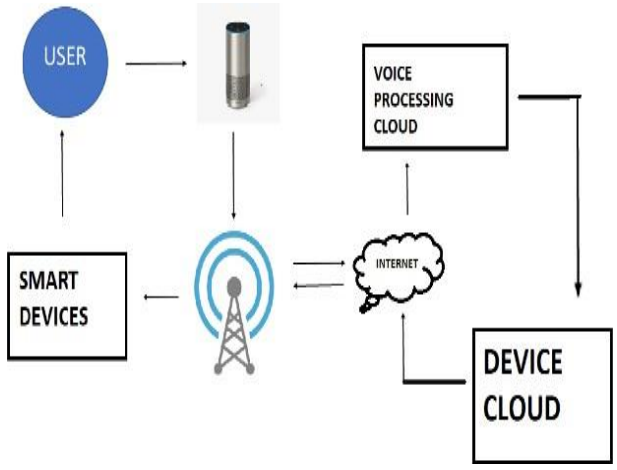

Fig.1 Working Model of ALEXA

Being such an important part of our day to day life, we should not neglect concerns for our security and privacy, everything presents in AI, machine learning and IOT can be hacked easily.

Also, the ALEXA, the most famous AI device has its various disadvantages which can't be ignored as they can cause a great loss to our privacy and two proof-of-concept attack (i.e. fake orders and home bulgur).

II. LITERATURE REVIEW

Katrina Rink et.al. [1] observed the interaction between Alexa and the user, the task that it can perform, and also the type of factors that can affect the working of Alexa behavior. In their report they collected the information from 19 participants through online interrogation and came to the

conclusion that it is mostly used by the people to know whether details, listening music, setting timers etc.

Those people who know the advanced features of this AI-powered machine used it 4-5 times more than the normal users. They found that the users are more satisfied with the way of interaction of these devices rather than getting the optimal results of their query.

Some of the people have managed to make emotional attachments with the device. But some of the users were not satisfied with some of the skills of Alexa like setting the timer as it is still under development and not that efficient, and also some children were not happy with its skills of making jokes because of the inappropriate jokes not relevant to their age. In the end, they concluded that still there is more work to be done on the Alexa so that it can overcome its drawbacks.

Xinyu Lei et.al. [2] suggested the 3 main problems of Alexa that are insecurity, privacy control of Alexa behavior and services. They also reveal the diverse effect of two proof-of- concept attacks.

In this, anyone can easily control any Alexa device to perform a various task like opening the room lights, doors, and do shopping from other's Amazon account.

They found that Alexa device can work without any human presence or without any nearby human activities because it lacks in human detection.

There is no voice biometric in it that means Alexa can't recognize its owner and if anyone speaks in more than 60dB sound pressure level" Alexa" can operate it.

The text to speech application converts the text document into speech. If there is any Bluetooth speaker near Alexa and got successful in connecting by fraud, by using this application he/she can operate Alexa easily.

The main aim of this paper was to tell about the lack of human presence detection. To overcome this, they proposed the concept of Virtual Security (VS) button with Alexa.

VS Button works on the principle of COTS Wi-Fi infrastructure that can be deployed nearby Alexa to detect tiny human motion (like walking around slowly). Once the machine detects the motion, they can activate Alexa device to take voice command.

They found that the Alexa works on a single factor authentication i.e. only the voice command is needed, no matter where it is coming.

Hyungi Chung et.al. [3] proposed the level of trust of these devices. They used the "Intelligent Virtual Assistant" word to refer to these devices. According to their statistics, they believe that the market of these devices will reach up to $21 billion until 2020.

They also suggested that all the IVA devices work as a cloud service which processes voice data. According to this, they classified IVA devices into two parts.

1) Built-in IVS (Siri, Cortana), and
2) Stand-alone IVS (Alexa, Google Mini).

In this we store our command as data in the cloud which is vulnerable and insecure, that can be used by the unauthorized entity for various frauds. They also examined DDoS (distributed denial-of-service) attack against Dyn LLC which exploited millions of home embedded devices, because of the same type of gateway used in IVS devices the possibilities of attacks may occur.

Qifan Pu et.al. [4] proposed the concept of WiSee, a gesture recognition sensor which is based on wireless signals. As wireless signals do not require any line-of-sight they can easily travel through walls and hard objects.

They made a proof-of-concept prototype, which works on the principle of USRP-N210s. The first commercially available as a gesture-based interaction was Xbox Kinect.

In their experiment, they found that can classify nine different types of gesture set with a total accuracy of 94%. WiSee can enable the whole-home gesture recognition system by using a few signal sources. We do not want a lot of signal sources for creating this.

Their experiment changed a Wi-Fi system into a gesture recognition system. To perform their experiment, they used two areas- an office environment and a two-bedroom apartment.

They worked over two main challenges that occur while changing a Wi-Fi into a gesture recognition sensor

1. How one can collect information about the different types of gestures from the wireless signals.
2. How they will deal with other human activities that are present in the environment.

Yuan Gong et.al. [5] examined a survey of recent attacks performed on a voice-controlled system like Alexa, google mini etc. and also some defense techniques to overcome these attacks. They proposed a classification of these techniques.

Voice Controlled System (VCS's) are the new ways of interaction with the machine, but they also give rise to a lot more risk concerns to our privacy and security. They can lead to can various attacks like replay attacks-in which one can replay a previously recorded sound commands to make a specific work done like opening a door, making an order and other types of tasks also some kind of self-triggered attacks.

The current defense technique which is present can only detect a specific kind of categories which only assumes that defenders already knows the details of attacking technology, which is not satisfactory.

Voice attacks can also be very hard to detect and can also be hidden with other types of sounds or by audio and video recorded files.

They classified the attacks which are based on implementation as-

1. Voice attacks which are based on the replay of commands.
2. Attacks which are on the operating system level based.

3. Attacks which are based on a hardware level.
4. Attacks which are based on the Machine level.

The existing defense technology is only able to solve only some categories; therefore, we require more powerful defense technique to protect voice-based AI devices.

Chan Zhen Yue et.al. [6] proposed model which extends the properties of the devices that supports Alexa Voice Services and also to create a cost-efficient, microprocessor-based smart home assistance device. Using their model, one can develop these devices according to their needs. To verify the model, they created a small room using a box and breadbox and connected its handset to the raspberry pi. They used a raspberry pi and Alexa skill kit to control these devices without any direct interaction.

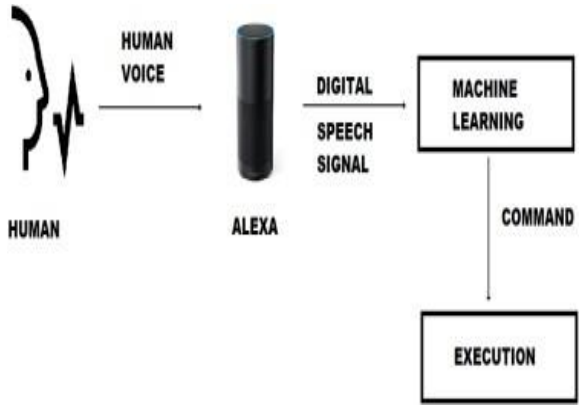

Fig. 2 Basic Voice Recognition System

According to them, after resolving the drawbacks and fraud risks from these devices we can use these devices as small healthcare system using Alexa services.

Tavish Vaidya et.al. [7] examined the gap which is present in the mechanism of speech recognition which is between human beings and the machines, as this difference can easily be exploited by any defaulter to produce sounds that are decodable as a command to a machine speech recognition system present in it.

They discussed how a wide range of these devices are affected to such changes and also described how a fraud attacker can use them to attack a victim.

Speech recognition is a very large complex problem which includes many of the techniques which use heavily biological terms like deep learning.

They also described a methodology to covert a speech to a form that can be understood by an MFCC's based systems.

MFCC (Mel-Frequency Cepstral Coefficient) represents the short-term power spectrum audios which are present on a non-linear frequency scale. Most of the voice recognition devices use MFCC's to show accounting features of provided input audio signals.

They presented a proof-of-concept attack that exploits the main difference between the recognition behavior of computers and human's speech.

III. FINDING AND DISCUSSION

The main aim of this paper is to highlight the main drawbacks, and the security and privacy terms so that one can feel safe while using these devices. These devices help our society in various terms but if handed in wrong hands can cause a lot of miss-happening to personal privacy and loss of money to many peoples. These devices are the future of AI and IOT, where one can easily operate and handle all his basic requirements just by uttering a few words. Days are not far when machines will be the best friends of humans, but as they are machines; they all have their pros and cons which one should never neglect.

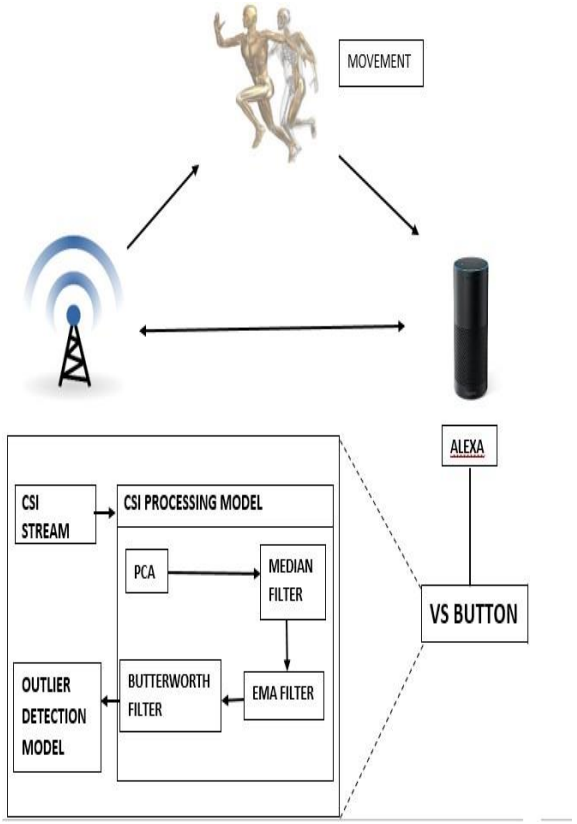

Fig. 3 Alexa with VS Button

We showed that how voice biometric[8,9,10] or human presence detection devices like VS buttons can be used to overcome the main drawbacks, which will further help to create those Alexa devices which are human's best friend not their enemy in any of the cases. There is a lot more work needs to be done on these devices and Alexa needs more skills (software programs in term of Amazon) to overcome its disadvantages.

By removing these drawbacks, we can use Alexa services in various fields of life rather than using it only in homes. They can be used in hospitals to heal patients with their excellent abilities to make interaction. They can also be used as a care taker of many places like old age home where its timer can be used to tell patients about their time of medicines. By

developing voice biometric we can make Alexa recognize his/her owner.

TABLE I. COMPARISON TABLE

Functionality	Amazon Alexa	Bixby	Siri	Google Assistant
Trigger	Hot-Word ("Alexa")	Dedicated button	Hot- Word ("Hey Siri")	Hot-Word ("Ok Google")
Voice Search	YES	YES	YES	YES
Text Search	NO	YES	YES	YES
Camera Search	NO	YES	YES	YES
Contextual Info	YES	YES	YES	YES
Reminders	YES	YES	YES	YES
Smart Home	YES	YES	YES	YES
Info Cards	YES	YES	NO	YES
Third-Party apps	YES	Supported	YES	YES
Languages	3	2(more on the way)	21	40(Voice Search) 5(Assistant voice)

IV. CONCLUSION

In the coming future, HDVA's will be an essential part of everyone's life. This paper presented the drawbacks of the Alexa devices and some solutions to overcome these problems. By making the voice biometric, or by making a human motion detection button we can make Alexa devices. We tried to study all the possible concerns that affect the behavior of Alexa devices and provide various methodologies to prevent these attacks as they can cause a great effect if misused.

REFERENCES

[1] Lopatovska, I., Rink, K., Knight, I., Raines, K., Cosenza, K., Williams, H., ... & Martinez, A. (2018). Talk to me: Exploring user interactions with the Amazon Alexa. Journal of Librarianship and Information Science, 0961000618759414

[2] Lei, X., Tu, G. H., Liu, A. X., Li, C. Y., & Xie, T. (2017). The insecurity of home digital voice assistants-amazon alexa as a case study. arXiv preprint arXiv:1712.03327.

[3] Chung, H., & Lee, S. (2018). Intelligent Virtual Assistant knows Your Life. arXiv preprint arXiv:1803.00466.

[4] Bhatia, S., Bajaj, J., & Roja, M. M. (2014). Technology, Systems and Implementation of a Smart Home Automation System: A Review. Suraj Bhatia Et Al, Int. J. Computer Technology & Applications, 5(5), 1690-1695.

[5] Gong, Y., & Poellabauer, C. (2018). An overview of vulnerabilities of voice controlled systems. arXiv preprint arXiv:1803.09156.

[6] Yue, C. Z., & Ping, S. (2017, April). Voice activated smart home design and implementation. In 2017 2nd International Conference on Frontiers of Sensors Technologies (ICFST) (pp. 489-492). IEEE.

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

[8] A. Das, O. K. Manyam, M. Tapaswi, and V. Taranalli, "Multilingual spoken-password based user authentication in emerging economies using cellular phone networks," in SLT, 2008.

[9] M. Kunz, K. Kasper, H. Reininger, M. Mobius, and J. Ohms, "Contin- ¨uous speaker verification in realtime." in BIOSIG, 2011.

[10] M. Baloul, E. Cherrier, and C. Rosenberger, "Challenge-based speaker recognition for mobile authentication," in BIOSIG, 2012.