Report for summer school CPS2018

LINGYUAN YANG
linyan@kth.se

July 30, 2018

Executive summary

This report is for the cyber security and privacy summer school 2018 in Trento. The name of our project is EspioNo. We provide device to secure your private conversation from listening by smart speakers. With a state-of-art technology disturbs the microphone but not human ear. To introduce our project, this report is divided into 4 sections. The first section will be focusing on the process of idea generation and optimization, as well as the technical basement. The second section will be introducing our business canvas, detailize our business structure. The third section will introduce our time line and expected cash flow. And the last section will evaluate the performance of myself both as a participant of summer school and a member of a team.

Contents

1 Problem and solution

1.1 Problem addressment

From the point that Amazon post its 90 seconds 'Alexa Loses Her Voice' advertisment, people begin to fascinate with smart speakers for the convenience it introduce into our life. However, to interacte with the users, smart speaker has to keep listening to them. This feature arise doubt and panic on privacy issue. Some reports publised recently seems to confirm that such concern make sense. According to The New York Times [1] and Quartz [2], alexa sometimes do record undesired conversation and even send it to other people. If this happens, why can't they tap us and make money by our privacy? Such tapping problem is a big worry for people who care about privacy. And the patent of Google and Amazon It's also been heard that when people talk about something they like or they want, an relative advertisment will be forwarded to them soon. A study has been done across America showing that 54% of the adults of the United States do not own a smart device in their home. 6.4 milion people said specifically they would not buy a smart speaker due to privacy reasons. 6.4 million people. And what about the 20 million people who want to increase their privacy. In the united states alone, we have not even talked about the rest of the world. And this problem will only increase in the future.
In this case, the problem we are going to solve is that people want to keep privacy while using smart speaker without undermine the convenience.

1.2 Problem validation

We use two ways to validate our problem. The first one is talking with people. The second way is reading reports and statistics about such problem. Firstly, we talk to people who has a smart speaker, 4 of our relatives are involved into our interview and all of them confirm that they do have such concern and they are willing to pay for products solving such trouble. We also happen to have "interivews" with Mr Jovan senior and junior during technical consulation and montoring slot. They both think Amazon and Google are not trustworthy and would like to have solutions for such problem, they represent the group of people who show interest about smart speaker but don't buy them due to privacy concerns.
We also collect statistics to validate our problem. According to a research conducted by Pew Research Center, 46% of the American adults have ever use digital voice assistants while 8% say they have ever use a stand-alone device such as Amazon Echo or Google Home. In the mean while, according to a survey from Business Insider [3] 40% of people are concerned about connected-home devices tracking their usage. And according to Business Wire , 6.4 million people said specifically they would not buy a smart speaker due to privacy reasons.

1.3 Solution and validation

1.4 Proposed solutions

To solve the addressed problem, we brainstorm and come up with several solutions and validate them one by one. Generally speaking, there are three ways preventing smart speaker from tapping.

- **Cut off power supply when not used** This way is to make a smart switch with speech recognition function to shut down the smart speaker when you don't want to use it anymore.

- **Monitor/Control the internet connection when not used** This way is to stand in between remote server and local speaker monitor all the internt traffic to secure no sensitive data been transmitted.

- **Physically prevent smart assistants from listening** This way is to disturb or deafen the microphone on smart speaker, make them unable to heard from us. And remove the interference when desired.

1.5 Validation process

In the beginning, these three solution all looks hopeful. But two of them are deprecated after we reading technical literature and doing test.

- **Why we abandon power switch solution** This solution is easy to validate. There is no doubt that it will work. But the problem is how well it works and if it's convenient. We do have a google home to conduct experiment. After we turn it off and reboot, it takes about 1 minute to be responsive again. Which means if we are going to use such solution, every time our customer want to use smart assistant they will have to wait for 1 minute after they say the key word.
  Customer buy smart speaker for the convenience. If our product makes life harder, it does't make sense and nobody will pay for it. In this case, we kill this solution first, even before look into technical detail.

- **Internet monitor solution** This solution actually include two different solution, internet switch and package sniffing. The second one is the inproved version of first one.
  The internet switch is what we originally come up with during brain storm. It works just like power switch, but instead of control over power, it controls upload traffic, prevent package being transmitted when not desired. The best thing about this solution is that it eliminates the drawback of booting latency. However, as long as we don't know the mechanism behind, we cannot ensure that the tapped audio won't be transmitted once reconnected to the internet.
  With such problem, we consult Yvo Desmedt during a consulation slot we booked. He suggest us to use sniffing technology to monitor what is actually transmitted. In other words, our product will act as a transfer center of package from smart speaker and analize the package to figure out if undesired package is being transmitted. With this suggestion, we do research and find out the audio data being transmitted are encrypted [4], which make sniffing work extremly hard. And it turns out somebody has already do simmilar job and we can derectly refer to it. According to a report [5] by Maik Morgensternfrom AV-TEST, the communication channel of Amazon Alexa uses TLS1.2 encryption with certificate validation/pinning. This technique prevented them to use a Man-In-The-Middle-proxy and read along the encrypted traffic. So they creat a scenario to evaluate the intensity of data transimission and estimate what is being transmitted. The scenario strats with 8 seconds of silence and then Asking Alexa "What time is it?" (6 seconds), then another 8 seconds of silence and followed by "Alexa – tea, earl grey, hot" and await the answer (5 seconds) finally the scenario end with 23 seconds of a normal conversation between two people (not including any Alexa keywords).

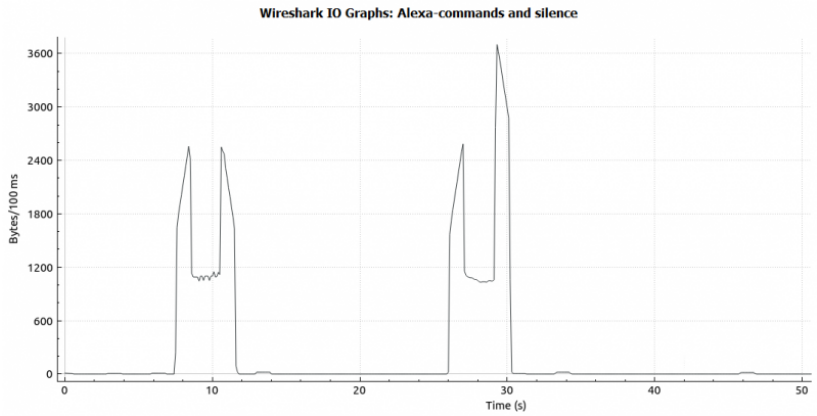

Figure 1: Transmitted bytes in the given scenario

We can see the result from figure 1 that the data transimission during conversation looks simmilar with silent time, it is very likely that Alexa do not send data if it's not working. This analisis is unfavorable to our solution.

Also, Prof. Yvo Desmedt give us an artical [6] about homomorphic encryption. With such technology we will be able to processing data, such as speech recognition, on remote server encryptedly wothout let the server konw a peice of information. To make use of such advanced technology we need to build our own smart speaker and the remote server to conduct homomorphic encrypted speech recognition. These is not viable for starting point, we plan to include it in our long term plan.

- **Audio disturbing solution** Interference is the first solution we come up with. The way how it works is to decrease the Signal Noise Ratio(SNR), which describe the level of a desired signal to the level of background noise, as low as possible. In other word, make noises to make conversation incomprehensive. There is one thing to be noted, a startup company called Privacy Shield has already

  – The first version of disturbing solution: use mini speakers to play white noise, which is a very common noise type, near the microphone of smart speaker. However, there is a big problem. The noise we play might bother our customer. According to the Acoustic Modeling report of google home [7], it could work well at SNRs ranging from 0 dB to 20 dB or above. So in the most optimistic scenario, we need to decrease the SNR to 0 dB which means the noise is as loud as human voice when reaching the microphone. This is load enough to be haerd several meters away.

  – The second version of our solution: use a case to cover the smart speaker which could both decrease the volume of human voice reaching microphone and the volume of noise needed. A case could fix the problem in the first version almost perfectly, it both decrease the volume of noise we need and the volem of noise leak out. But we need a mechanical structure to open and close the case, which should also be audio-controled. The way it work is shown in figure 2. This structure is viable although it require complicated mechanical structure to support smooth opening and closing of the case and also speech recognition function to remain the user experience. Not good enough but we still use this solution for our first pitch.

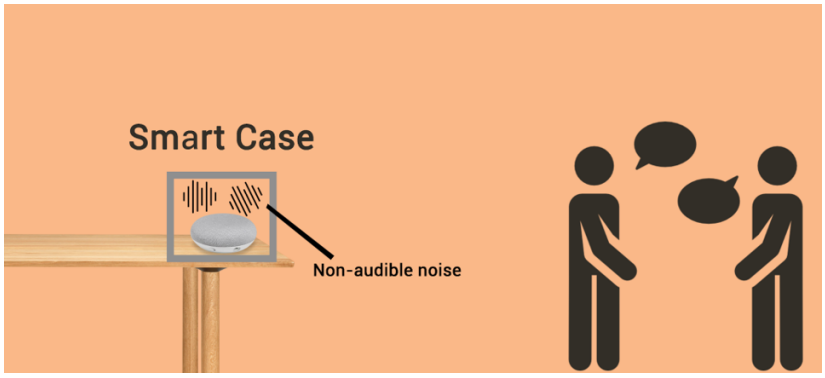

Figure 2: Audio disturbing solution

1.6 Final solution

The final solution actually evolved from the second version of Audio disturbing solution. After struggling on simplify our product and trying to add encryption elements to the solution. Fortunately, we found a newly published artical from UIUC [8] which provide a way to use inaudible ultrasonic sound to disturb microphones.
Inside the microphone, all the physical audio will first been transfered to electronic signal, after that the signal should pass a component called pre-amplifier which will amplify the analog signal to make sure it can be effectively measured by ADC. In this procedure, the output of pre-amplifier is linear only when input is in the audible frequency range which is 10-23 KHz, outside this range, the response exhibits

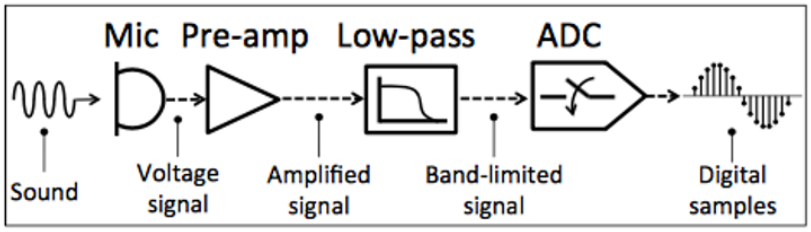

Figure 3: Circuit inside microphone

non-linearity which will introduce harmonic. In the artical, by making use of such non-linearity, they design the input ultrasonic sound and make a controlable shadow noise signal of 10KHz, which has been perfectly record by microphone.

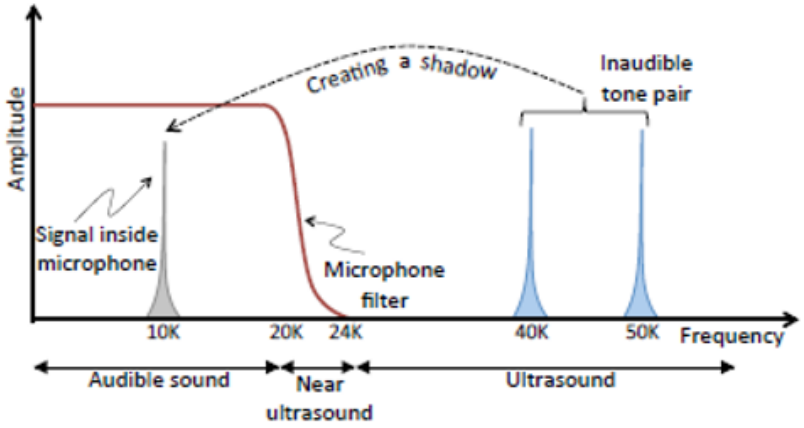

Figure 4: Use ultrasound to generate shadow noise

2 Business modelling and planning

2.1 Business modelling

2.2 Business planning

3 Business development process

4 Self evaluation

References

[1] S. Maheshwari, “Hey, alexa, what can you hear?” https://www.nytimes.com/2018/03/31/business/media/amazon-google-privacy-digital-assistants.html, accessed July 5, 2018.

[2] quartz, “An oregon family’s encounter with amazon alexa exposes the privacy problem of smart home devices,” https://qz.com/1288743/amazon-alexa-echo-spying-on-users-raises-a-data-privacy-problem/, accessed July 5, 2018.

[3] businessinsider, “Consumers are holding off on buying smart-home gadgets thanks to security and privacy fears,” http://www.businessinsider.com/

consumers-holding-off-on-smart-home-gadgets-thanks-to-privacy-fears-2017-11?IR=T/, accessed July 5, 2018.

[4] Google, "Data security and privacy on google home," https://support.google.com/googlehome/answer/7072285?hl=en, accessed July 5, 2018.

[5] M. Morgenstern, "Careless whisper: Does amazon echo send data in silent mode?" https://www.iot-tests.org/2017/06/careless-whisper-does-amazon-echo-send-data-in-silent-mode/, accessed July 5, 2018.

[6] F. Bourse, M. Minelli, M. Minihold, and P. Paillier, "Fast homomorphic evaluation of deep discretized neural networks," Cryptology ePrint Archive, Report 2017/1114, Tech. Rep., 2017.

[7] B. Li, T. Sainath, A. Narayanan, J. Caroselli, M. Bacchiani, A. Misra, I. Shafran, H. Sak, G. Pundak, K. Chin, K. C. Sim, R. J. Weiss, K. Wilson, E. Variani, C. Kim, O. Siohan, M. Weintraub, E. McDermott, R. Rose, and M. Shannon, "Acoustic modeling for google home," 2017. [Online]. Available: http://www.cs.cmu.edu/~chanwook/MyPapers/b_li_interspeech_2017.pdf

[8] N. Roy, H. Hassanieh, and R. Roy Choudhury, "Backdoor: Making microphones hear inaudible sounds," in *Proceedings of the 15th Annual International Conference on mobile systems, applications, and services*, ser. MobiSys '17. ACM, June 2017. ISBN 9781450349284 pp. 2–14.