Real time Alexa packets profiling analysis

Giuseppe Polese
gpolese@unisa.it
Universitá degli studi di Salerno
Salerno, Italy

Bernardo Breve
Stefano Cirillo
bbreve@unisa.it
scirillo@unisa.it
Universitá degli studi di Salerno
Salerno, Italy

Biagio Boi
b.boi@studenti.unisa.it
Universitá degli studi di Salerno
Salerno, Italy

ABSTRACT

Nowadays, the introduction of home virtual assistents like Alexa Echo or Google Home became a practice, just considering that over 27% of families owns one.

It's obvious that those devices simplified the life by creating a smart house with few money; but what's the impact these devices have on people privacy? There are a lot of cases in the United States in which the judge asked to Amazon to provide the recording done by the Echo Dot in order to find helpful evidences for the case; so, the question is: "It's possible to prevent the sending of sensible informations to the servers when the weak word is not pronunced?"

In this project we will profile each packet exchanged between the Alexa Echo and the Server in order to classify the nature of the packets and consequentially we will use a machine learning model to discover when the Echo Dot is sending an inappropriate packet.

KEYWORDS

data analytics, alexa, packets profiling

1 INTRODUCTION

TODO

2 ALEXA ARCHITECTURE & SECURITY

The Alexa architecture isn't really easy to explain, we will resume just the keypoint in order to better understand the main functionalities for our purpose.

(1) Alexa is always in listening waiting for the weak word to be pronunced to start the recording of the voice;
(2) From the weak word, till the end of commands, Alexa will record the speech and partially sends it to Alexa Voice Service, that can be considered as the brain of Alexa;
(3) Alexa Voice Service will process the audio using Natural Language Processing and Natural Language Understanding in order to retrieve a response for the given request.
    (a) Natural Language Processing (NLP) improve the Word Segmentation that separate a chunk of continuous text into separate words.
    (b) Natural Language Understanding (NLU) is a subtopic of NLP and uses the AI to map text to the meaning[1] in order to understand the speech and the request.
(4) Depending on the sent command, the Voice Service will take an action (turn on the light) or send the information back to the device and Alexa may speech.

3 ALEXA FLOW OF COMMUNICATION

TODO

4 PACKETS ANALYSIS

TODO

5 DATASET

TODO

6 MACHINE LEARNING

TODO

7 CONCLUSIONS

TODO

REFERENCES

[1] Paul Semaan. 2012. Natural Language Generation: An Overview. *Journal of Computer Science & Research (JCSCR)* 1, 3 (2012), 50–57.