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 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 is sending an inappropriate packet.

KEYWORDS

data analytics, alexa, packets profiling

1 INTRODUCTION

The evolution of smart devices over last years has increased exponentially and the introduction of these devices within the house is progressively growing. The major problem related to these devices is that usually the privacy is not considered, although there are a lot of regulations (just see the GDPR) that describe how the user data have to be stored and who can access to these data. Starting from these two points I decided to understand what happens within the context of smart assistance device such as Alexa Echo. Reading on the Amazon Alexa website it's clear how and when they send data to the Amazon Cloud, but in opposition we have some cases in which the recording done by the Echo have been used during the processes in order to extract some evidence. In order to clarify this aspect and the role of the smart assistance devices I decided to analyze the outgoing traffic from the device and classify it by considering the environmental conditions. The first phase of this analysis consist in understand which kind of packets are sent from the device in order to create a complete dataset of packets sent by the device; during the second phase I try to understand if exist a pattern that identify the packets that souldn't be sent to the Cloud by creating a machine learning model able to identify them.

2 STATE OF ART

Wirte here all papers found related to the project

3 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 wake 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[**?**] 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.

All the communications between the Echo and the Server are secured by using an SSL/TLS encryption schema. In particular, the Echo looks for an available Amazon Server each 3 minutes and then establish a connection (by following the SSL Handshake). Once the communication has been initialized it starts to send different packets of different size always encrypted; clearly, these pakcets may contains both authorized and unauthorized data. It's important to underline that all the recordings are available on the Amazon platform and can be reproduced and deleted whenever the user wants only by accessing to the platform and this guarantee more transparency to the user that can directly handle all the recordings. The point is that here we found only the recordings done by the Echo in an authorized way (when the wake word it's been pronounced), but this doesn't imply that some other recordings can exist and be stored in hidden way.

4 ALEXA FLOW OF COMMUNICATION

Describe here the entire flow of communication

5 PACKETS ANALYSIS

In order to create a good dataset we will classify the packets sent by the Echo. In particular is possible to classify the type of packet by

looking at flags, packet size and protocol used. There are different types of packet that will later be included in the dataset as classes:

(1) **handshake**: At the beginning of each new SSL/TLS communication between the Echo and the Server there are different packets exchanged in order to establish a secure communication. These packets can be easily identified by the protocol used for the communication (TLS) and by checking the flags related to the content of the message, in particular, the possible flags are:
   (a) Change Cipher (20).
   (b) Server Hello - Show Certificate - Encrypted Message (21).
   (c) Alert (22).
   Echo changes communication Server very frequently, about each 2 minutes.
(2) **syn**: Two packets with fixed length are sent from the device in a fixed interval, for this reason we believe that these are synchronization packets. These packets may be sent to guarantee the synchronization with the Amazon Alexa application and with the Amazon servers. There are two possible synchronization packets:
   (a) A packet of 100 byte is sent each 30 seconds (30002 milliseconds)
   (b) A packet of 99 byte is sent each 90 seconds (90820 milliseconds)
   The communication always happens using SSL/TLS, so it's impossible to decrypt the content of these message and confirm that is securely a synchronization packet.
(3) **ack**: These are the classical packets used to confirm that the received packets are valid and successfully received.
(4) **retransmit**: These packets are used to retransmit the data that didn't received an ack, usually this happens when the Echo try to communicate with other Amazon devices (Fire Stick for ex.) but doesn't receive response.
(5) **app_data**: These packets are relative to the normal communication of the Echo and can be easily recognized since the communication happens over TLS/SSL protocol and by checking the flag of the considered packet that is equal to 23. In particular, we will include in this class all the communication packets that include also recording of voice occurred during the conversation between the Echo and the final costumer.
(6) **not_relevant**: We will include in this category all packets that don't match any of the other categories, and are not relevant for our study.

It's import to underline that the packets classified as app_data may include both authorized and not authorized packets. The unauthorized packets include all those packets sent in a context in which no application date should be sent to the Server by considering the environmental variables (no wake word pronounced) and hardware aspects (microphone muted by pressing the relative button); for this reason is important to introduce some features related to these aspects. The aim is to distinguish three kind of packets:

- **not_justified**: packets sent to the Server in a context in which the Echo is unable to send these data by considering the hardware aspects.
- **justified**: packets sent to the Server without any interaction between the Echo and the Customer, but the Echo is able to record.
- **expected**: packets sent to the Server since an interaction between the Echo and the Customer happened.

Clearly, the only allowed packets are those classified as "Expected".

6 DATASET

In order to create a valid and well-distributed dataset we analyzed the behavior of the Echo in different context, in particular by considering different cases:

(1) **Microphone mute**: when the hardware button is pressed and the device should be unable to send application data to the server.

Figure 1: Echo when the disable microphone button has been pressed.

(2) **No wake word pronounced**: when the device is able to listen the voice but shouldn't send to the Server any packets of application data.
(3) **Normal stream of data**: when a communication happens. It may include different questions or requests, for example:
   (a) How is the weather today?
   (b) When does Liverpool play?
   (c) Add water to the cart.
   All these questions or requests expect a response from the Server, so a huge amount of ack will be sent from the device.
(4) **Stream of data**: when the user request for a song and the Echo start to stream it.

7 MACHINE LEARNING

TODO

8 CONCLUSIONS

TODO