

Privacy Mediator for Audio Data

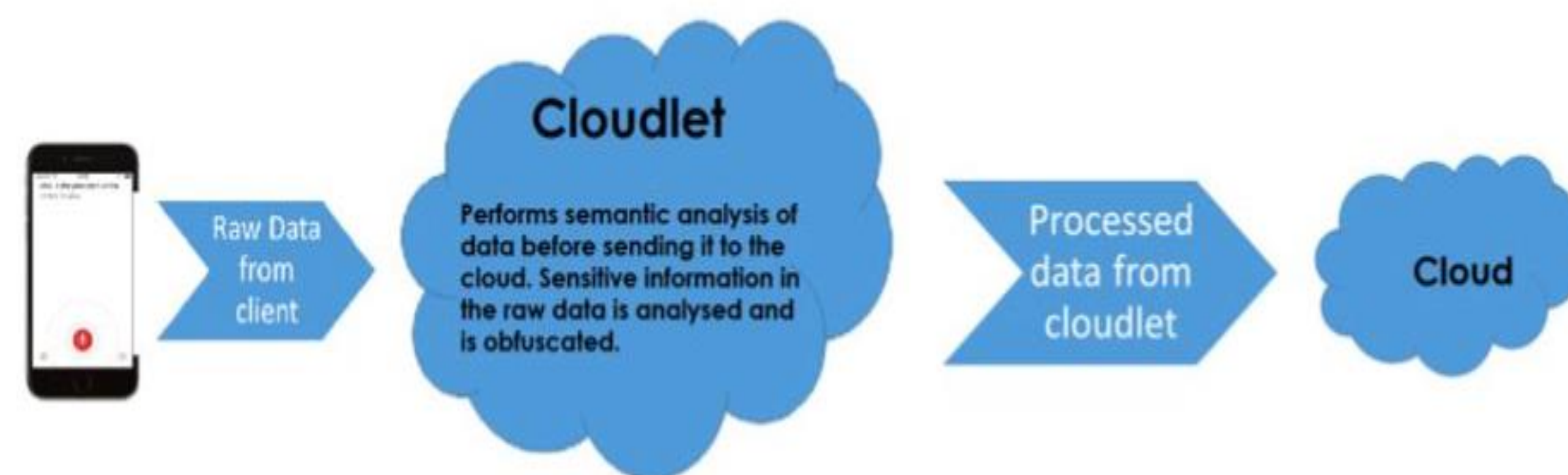
Mentor – Junjue Wang
Ayushi Singh, Rajat Pandey, Ankit Jain

Problem Statement

- No semantic analysis of user data prior to its transmission to cloud.
- Exposure of unprocessed audio data on cloud might be a reason for violation of user privacy.

Proposed Solution

- Raw data from user device is transmitted to cloudlet before sending to cloud.
- Cloudlet performs semantic analysis over the recorded voice data and obfuscate sensitive information like SSN, Credit cards.
- Generalizing the user language model and extracting user specific sensitive information.

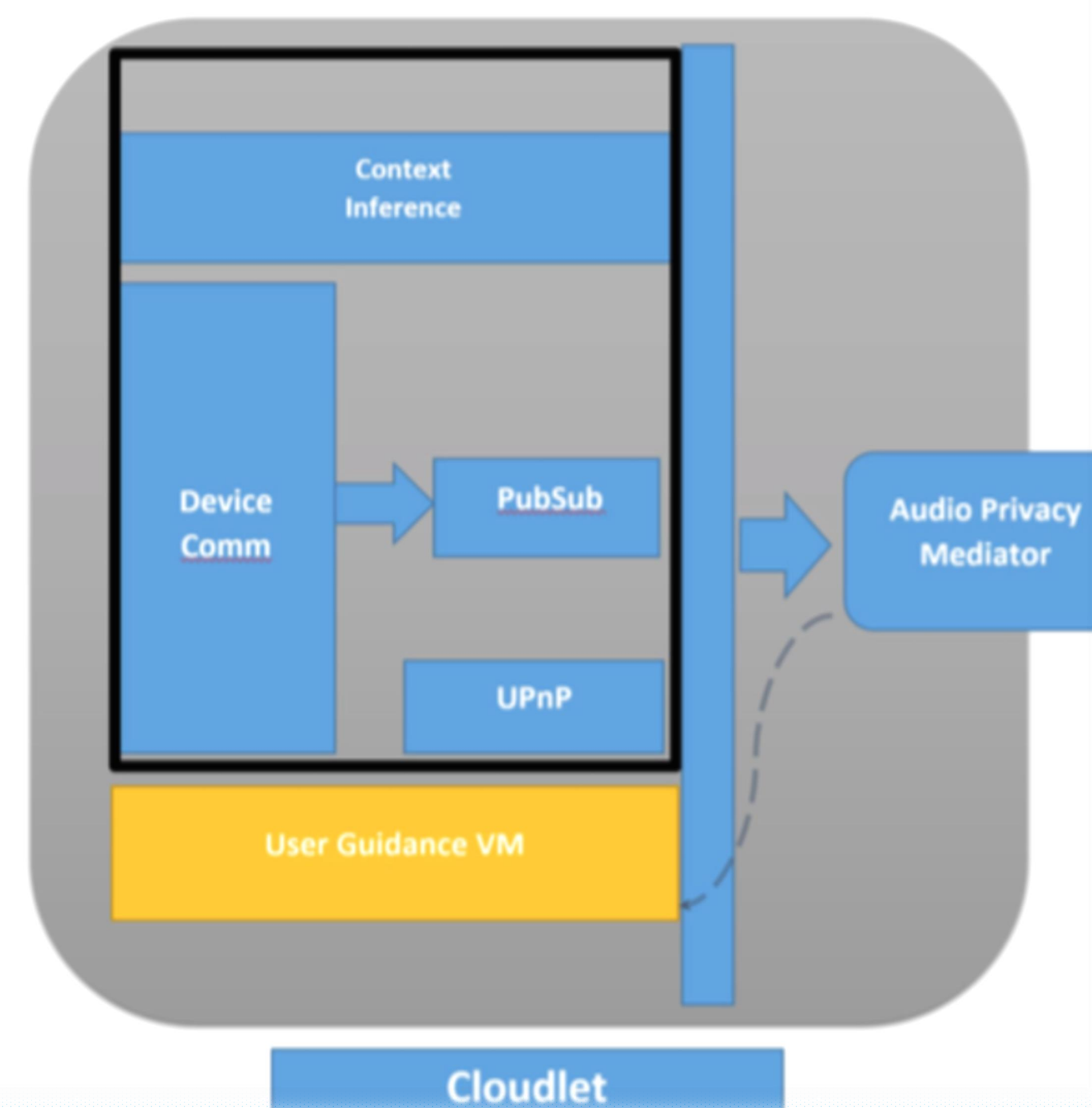


Technology Stack

- Gabriel – To transmit and receive the data from client device to cloudlet in the form of byte stream.
- Using CMUSphinx for conversion of text to speech and vice versa.
- Using Pattern Matching and Regular Expressions to spot and remove the sensitive information from recorded voice.

Architecture

Cloudlet Architecture



Device Comm

- This component is used to interact with the device, and acts as mediator for streaming user's information to and fro across cloudlet and the device.

PubSub

- This component publishes the stream of data to the corresponding subscribed modules.

User Guidance VM

- Outputs of Audio Privacy Mediator are sent to user guidance Vm that integrates the output and performs Semantic Analysis of recorded voice data.

Upnp

- VM at cloudlet discovers the sensor streams of interest through Upnp discovery mechanism.

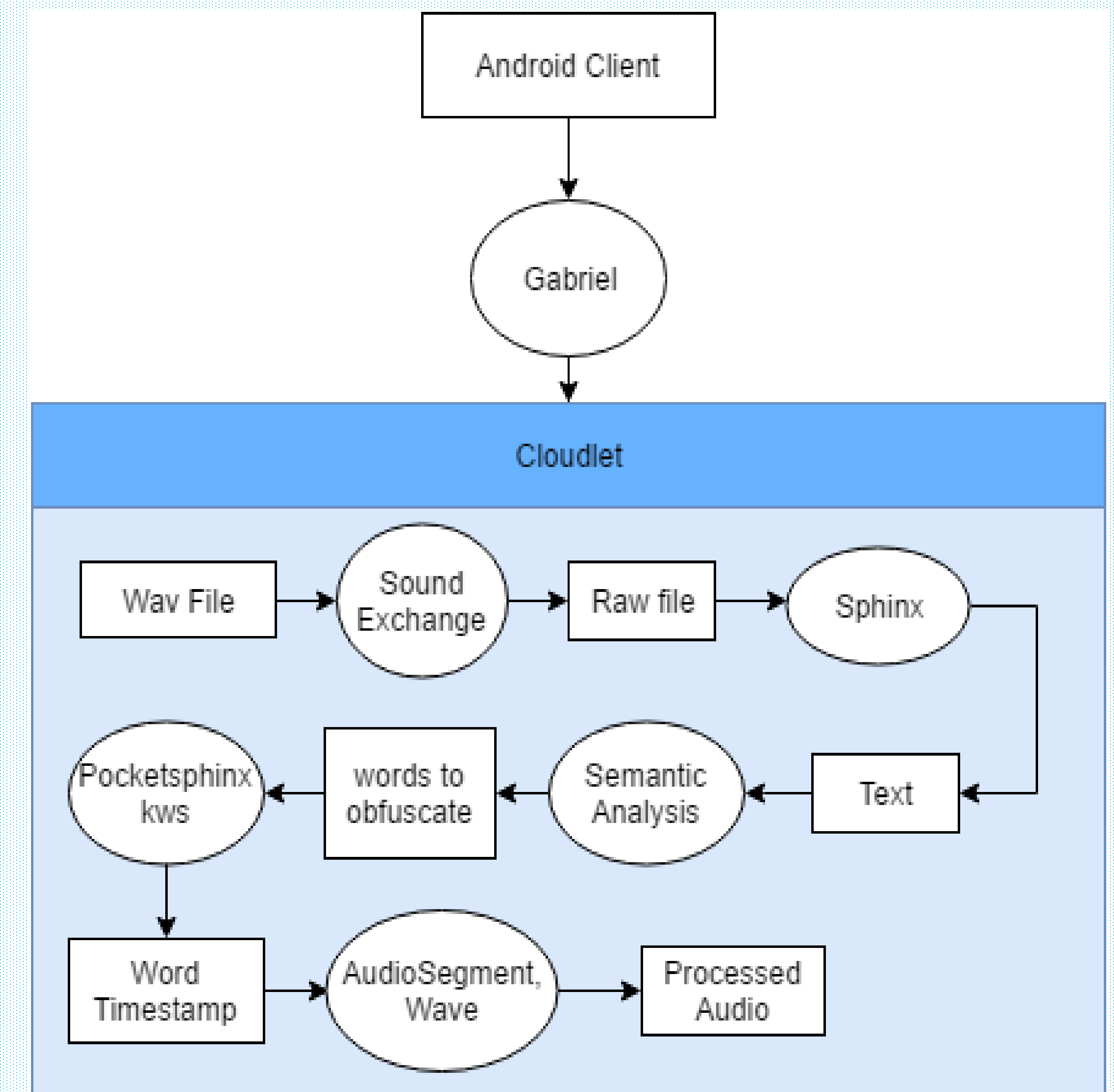
Context Inference

- Context- sensitive control of sensors on the device is achieved through a context inference module.

User Guidance VM

- Outputs of Audio Privacy Mediator are sent to user guidance Vm that integrates the output and performs Semantic Analysis of recorded voice data.

Implementation



Results

- Results can be better explained with the help of example:
- User speaks a sentence which has sensitive information:
- User : ".....my password is marshmallow."
- The sentence is recorded and sent to cloudlet. The cloudlet performs pattern matching against a number of pre defined patterns. Few of the patterns are :
- Sample Pattern 1: "(?<="+keyword+" is).*(?=\\)"
- Sample Pattern 2: "(?<="+keyword+").*(?=\\)"
- The matched password is then spotted and removed from the voice recording.