CECS 491A

Analysis Documentation

Khanh Nguyen (Leader)

Angel Franco

Christopher Imantaka

Ryan Valdriz

Table of Content

Purpose

This document is for the developers to explain the specification of the technology used in this project. The document will give a thorough description of our algorithms, protocol, required computer specification, database, third party applications, machine learning software, and integration. With all of the stated technology we will, discusses their utilization, possible point of failure, how they work, and performance evaluation.

Summary

This document goes into details on the specific technology, architures, softwares, api's, and protocols used to accomplish functionality, performance and security. The layout of the document will be divided into sections. The sections are divided up and follows the order of the table of content. Each sections will cover a different capability that are essential for our project. For each section, a description will be given to explain what will be used, why, and how. This document is meant to guide the developer on what is necessary to build and use this project as well as an evaluation of the mechanism suggested.

Video Chat and Text Chat

- Protocol
    - UDP for video streaming
    - TCP for Text Chat
- Requirements for video streaming:
    - WebRTC : Collection of communications protocols and APIs that enable real-time peer to peer connections within the browser.
        - Compatibility: Chrome, Firefox, Edge, Opera
    - Server
        - Signaling : Process of discovery connection end and exchange negotiation message for other WebRTC peers is handle through signaling server.
        - ICE Candidates : Two peers exchange ICE candidates until they find a method of communication that both supported.
        - STUN : The server used to make calls between two different networks.
        - TURN : Connect past proxy firewalls in networks.
- Advantages of UDP for video streams:
    - Less overhead
    - Scales well
    - Low latency
    - Good throughput
        - Relative to TCP
- Disadvantages of UDP for video streams:
    - UDP streams can have problems passing through firewalls and NATs
        - UDP hole punching may not be possible due to port randomization by the NAT
    - Possible drops in quality of the video-stream
- Advantages of TCP for text chat:
    - Packets arrive in order
    - Lost/dropped packets will be resent
- Disadvantages of TCP for text chat:
    - Slower transmission

For our video chat module of our app, we will be using udp. The reason for this is because the videostreaming should be in real-time. Live-video streams are somewhat fault-tolerant so even if some packets are dropped, content would still be displayed, just in a lower resolution. Using TCP would mean that lost/dropped packets would be resent. For live video streams, this would be bad, since only the newest packets are of any importance. Also, it would mean that newer packets would have to wait for older packets to be resent. This is why we will be using UDP for our streaming protocol. UDP does not care about lost/dropped packets. It will allow our video chats to be as up-to-date as possible.

How The User Gets Access to Our Application?

Our application will be a desktop app. The user will be able to install the desktop app onto their device by accessing it through our github. The user will simply need to use any web browser to reach our github. Our application will on be in a folder freely accessible for the user to download and install on their device. This software will be seen as open source.

What Are The Requirements For a Computer to Use Our Application?

After downloading our desktop app, the computer specifications will need to meet the requirements stated in this section. These requirements are essential for performance, proper functionality, and capability. The operating system needs to be Windows 10 or the latest mac OS. The reason why we require this operating system is because the platform that our software is developed on and for is mac OS and Windows 10. Since our app includes communication through video chat and the translation of American Sign Language(ASL) utilizes a webcam; we require the camera to have minimum resolution of 720p HD, video rate of 30 FPS, and standard lens. Anything but a standard lens and the the camera will disrupt the image processing task of the machine. Darkness or extreme brightness are light exposures that will harm the hand recognition performance of our software. The computational power required will include a CPU processor minimum requirement of 2 GHZ. Additional computational power includes a minimum graphic card performance to that of Intel HD graphics 4000 graphics with 384 MB of shared DDR3 memory. The minimum volatile memory must be 4 GB, and user will need approximately 500 MB of nonvolatile memory to permanently store the program on their computer.

Screen Reader Software

A screen reader is the interface between the computer's operating system, its applications, and the use. The screen reader software is responsible for reading the text on the screen out loud. The software is useful for people who have trouble reading text on the screen or just rather have the computer read the message out loud. For our project we decide to use NVDA (NonVisual Desktop Access) for windows. NVDA is a free software and is easily controlled by moving the cursor to the relevant area of text with a mouse or using the arrows on the keyboard. NVDA is open source program for developers like us to download and use.

Speech Recognition Software

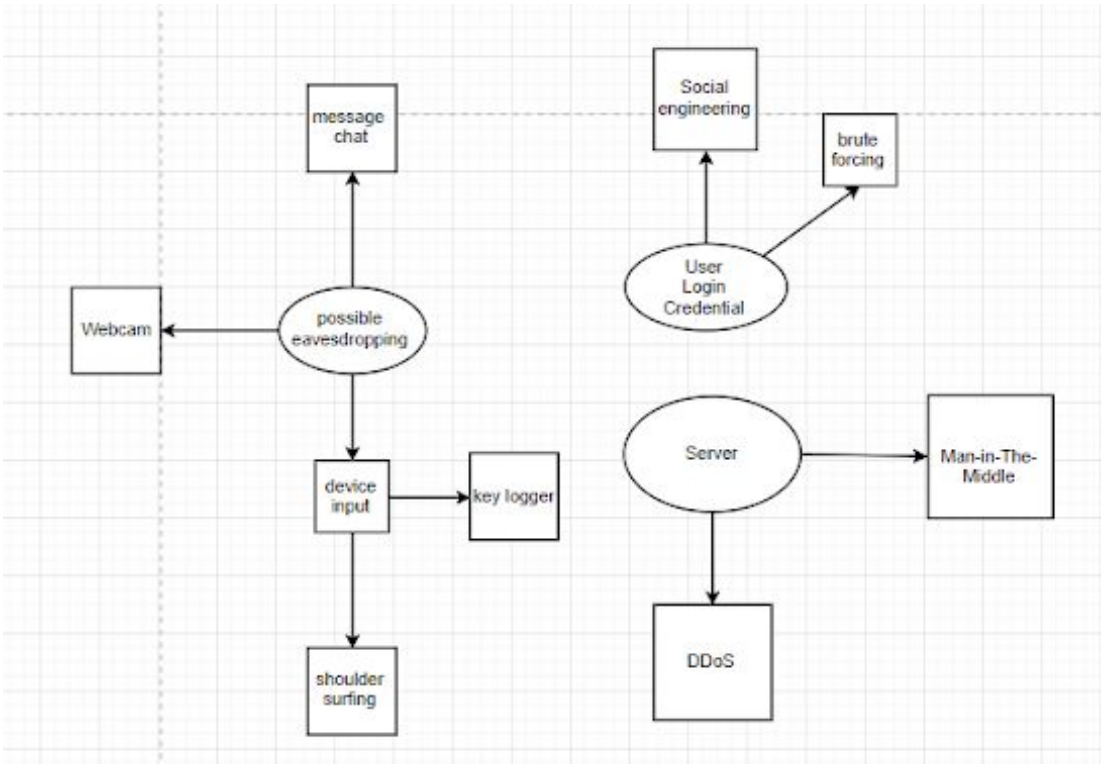
We need a speech to text software in case we have users who can not see or for user who want to speak while having a conversation rather than typing. This can also be useful when an ordinary user and a user with a hearing disability are having a conversation because the deaf person can not hear and needs subtitle to read what the other user has said. Amazon has an speech recognition API called Amazon Transcribe. I can receive a live audio stream and return a stream a transcript in real time. It can be used to generate subtitles or content which is what we want. Whats is great about this service is Amazon Transcribe is continuous learning and improving to keep up with the evolution of language. A evolving speech recognition software is crucial and an advantage over other softwares because language is constantly changing. Amazon Transcribe service is very inexpensive to use its service for free 60 minutes a month, Amazon Transcribe API is billed monthly at a rate of $0.0004 per second. This is accomplished

Security

- Authentication/Authorization
    - JWT(JSON Web Token)
        - When users login, we will grant them a JWT to use as authentication.
        - Users will use this JWT for authorization as well
        - Pros for using JWT:
            - Easy to scale horizontally
            - Easier to maintain and debug than using sessions
            - Useful for creating true RESTful services
            - Built-in expiration functionality
            - Self-Contained
        - Cons for using JWT:
            - Payload is in plaintext
                - If a jwt is captured, it's easy to read the payload.

- Confidentiality
    - The chat will be encrypted using AES and RSA encryption suites
        - AES
            - Time it takes to find a 256-bit key by brute force
                - Around 3×10^{51} years
- Integrity
    - The integrity of the messages will be assured by using a hash-based message authentication code(HMAC)
- Assets
    - User Information
- Stakeholders
    - Users
    - Programmers
- Adversary Models
    - For users-user information
        - Active outsider adversary
            - Digital resources
                - Low amounts of computational power
                - Power supply
                    - Limited resources
                    - Not worth spending too much power to obtain information
        - Passive outsider adverary
            - Digital resources
                - Low amounts of computational power

            - Power supply
                - Limited resources
                - Not worth spending too much power to obtain information
    - For programmers-user information
        - Active outsider adversary
            - Digital resources
                - Low/medium amounts of computational power
                - Power supply
                    - Limited resources
                    - Not worth spending too much power to obtain information
        - Passive outsider adverary
            - Digital resources
                - Low/medium amounts of computational power
                - Power supply
                    - Limited resources
                    - Not worth spending too much power to obtain information
        - Active insider adversary
            - Digital resources
                - Low amounts of computational power
                - Power supply
                    - Limited resources
                    - Not worth spending too much power to obtain information
                - Might have access to database
                - Might have access to source code
        - Passive insider adversary
            - Digital resources
                - Low amounts of computational power
                - Power supply
                    - Limited resources
                    - Not worth spending too much power to obtain information
                - Might have access to database
                - Might have access to source code
- Attack Surfaces

- 

- Counter Measures
  - HMAC
  - Encryption
    - AES
    - RSA
  - Salting passwords
  - Let’s Encrypt Certificate
  - TLS 1.2

Architecture

We are using 3D convolutional neural network as the learning algorithm for our software. Convolutions are filter (matrix / vectors) with learnable parameters that are used to extract low-dimensional features from an input data.They have the property to preserve the spatial or positional relationships between input data points.Convolutional neural networks also exploits the spatially-local correlation by enforcing a local connectivity pattern between neurons of adjacent layers. Intuitively, convolution is the step of applying the concept of sliding window (a filter with learnable weights) over the input and producing a weighted sum (of weights and input) as the output. The weighted sum is the feature space which is used as the input for the next layers.

We are using 3D convolutions which applies a 3 dimensional filter to the dataset and the filter moves 3-direction (x, y, z) to calculate the low level feature representations. Their output shape is a 3 dimensional volume space such as cube or cuboid. They are helpful in event detection in videos, 3D medical images etc. They are not limited to 3d space but can also be applied to 2D space inputs such as images. In order to implement this, we are going to use TensorFlow API to build this architecture and train the machine in order to create a model for the application. Result will be a calculated graph that use to compare to frames input from the webcam.