Multi-User Backend for Meeting Translation Project

Fahim Imaduddin Dalvi
fid@qatar.cmu.edu

Francisco Guzman
fguzman@qf.org.qa

Introduction

- The Meeting Translation project aims to provide:
  - A platform to make meetings involving multiple languages more efficient
  - Real time speech decoding
  - Machine translation
- Currently, when multiple languages are involved, human interpreters are brought in to bridge the gap between the meeting participants
  - This limits participation in the meeting as communication is complicated and not direct

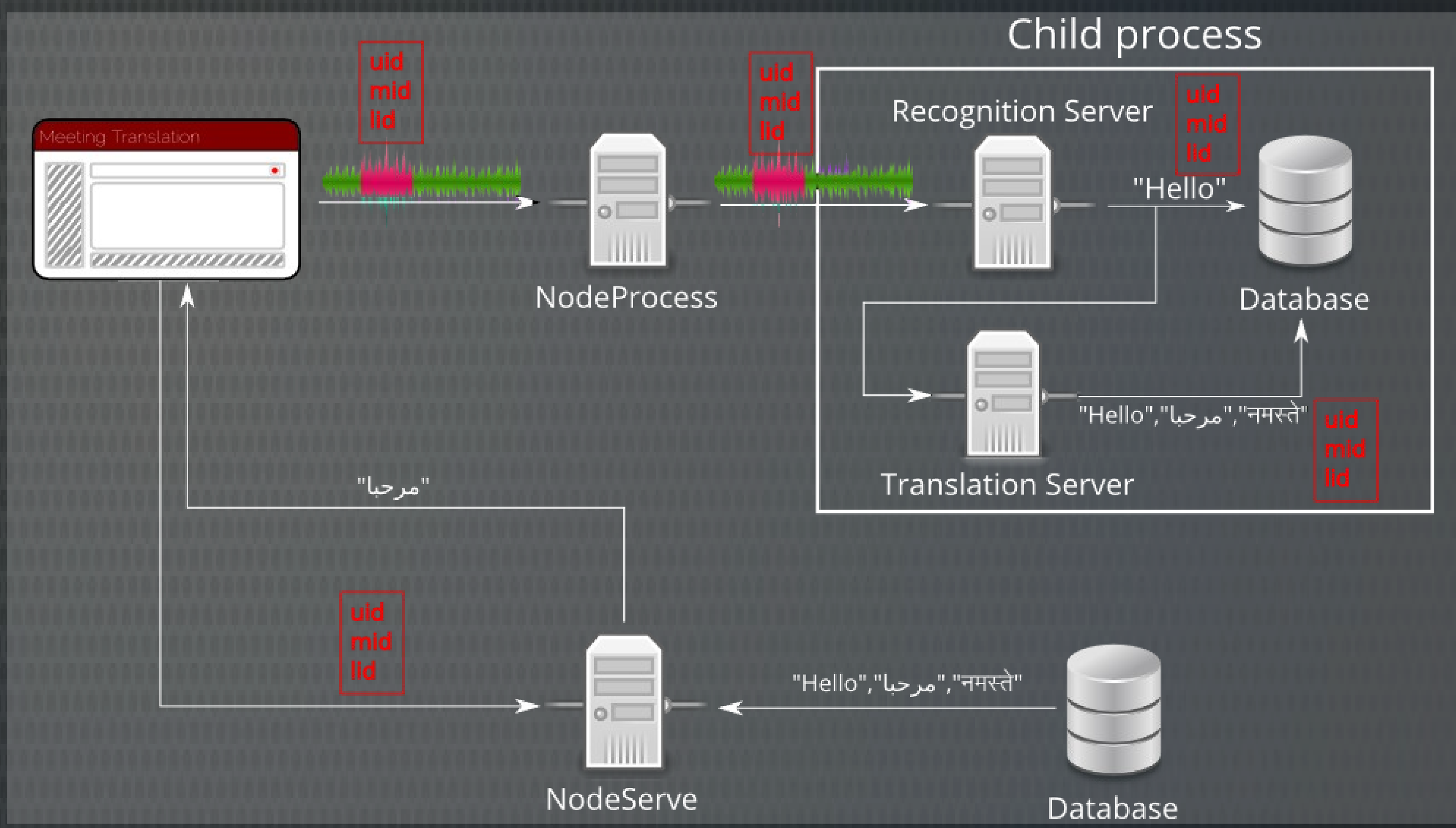
Problem

- Although we already have very efficient methods for performing real-time speech recognition and translation, we do not have a system that brings these technologies together in a way that consumers can conduct meetings in a seamless fashion
- There is a need for a system that can support multiple users and multiple meetings at the same time
- The problem that this project aims to solve is to build the backend for such a system, using new technologies such as WebRTC so that it is compatible with all newer generation devices

Solution

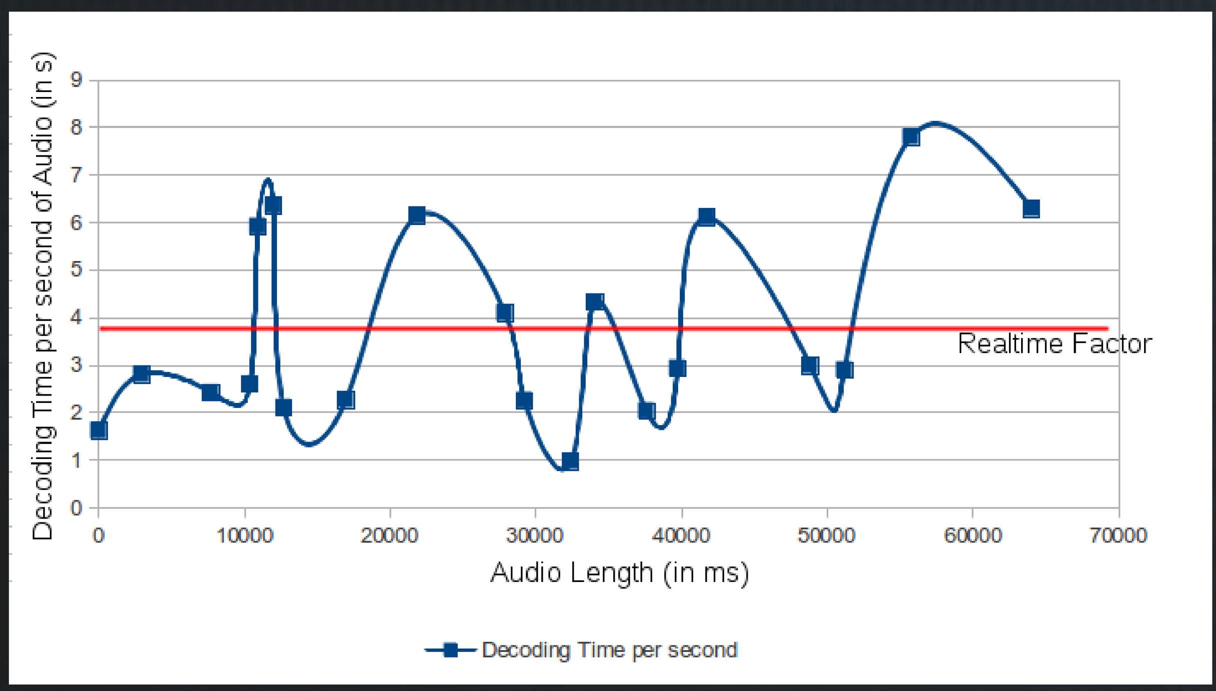
The proposed solution has the following key features:

- Multiplexed Architecture – Multi-user and Multi-Meeting
- Pipelined Architecture
- Built-in capability to collect statistics about the performance of each of the modules
- Robust and efficient – Based on the new and upcoming NodeJS and WebRTC technology
- Scalable – Currently supports two languages, but more can be added easily



Results

- The system was built completely and was tested thoroughly
- Multiple meetings with multiple users each were conducted simultaneously
- The system was also tested with meetings where both the Arabic and the English languages were involved
- The statistics collection service was also used to figure out where the bottlenecks were present in the system



Future Work

A few features that would improve the overall system may be:

- Continuous Audio – Currently, the audio is chopped into small chunks to get as close to real-time performance as possible. The accuracy of the system would improve considerably if we were to truly process continuous audio
- Integrate more languages – Although the system itself can easily accommodate more languages, the backend services for translation and speech recognition are not yet available

معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
Member of Qatar Foundation عضو في مؤسسة قطر