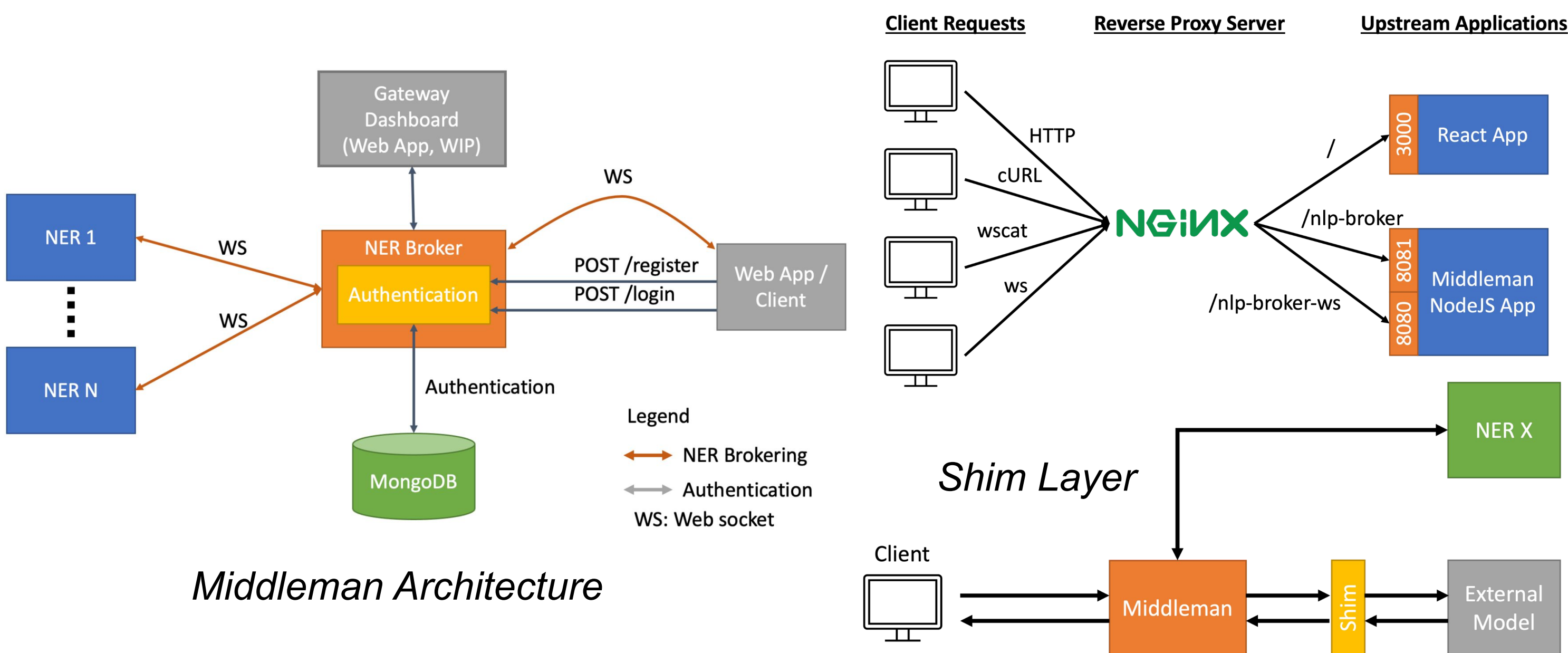


Named Entity Recognition Middleman Web Application

Low-Latency Real-Time Platform for running multiple NLP Models

Student: Kee Wan Ting Supervisor: A/P Chng Eng Siong



Middleman Architecture

Project Objectives:

Full stack web application to democratise access to NLP models by allowing developers to host their NLP models for public use. Web app uses HTTP RESTful APIs for registration and login, and stores users credentials in MongoDB.

Backend uses a hybrid one-to-many multiplexing WebSocket strategy for concurrent queries to the same NLP model. NGINX acts a reverse proxy server to process client requests and responses. Finally a shim layer that transparently intercepts API calls and update arguments on a per model basis enables easy on-boarding of external NLP models.

Integration and maintenance of the application is the top priority. The Middleman is extensible and supports multiple local and hosted NLP models to be seamlessly integrated, with minimal maintenance overhead.

Graphical User Interface (with e.g.)

