SUPER: Seamless Urban Pedaling – Efficient and Reliable

6.1800 Preliminary Report

Abhay Basireddy, Francisco Colon, Tae Wook Kim

March 22, 2024

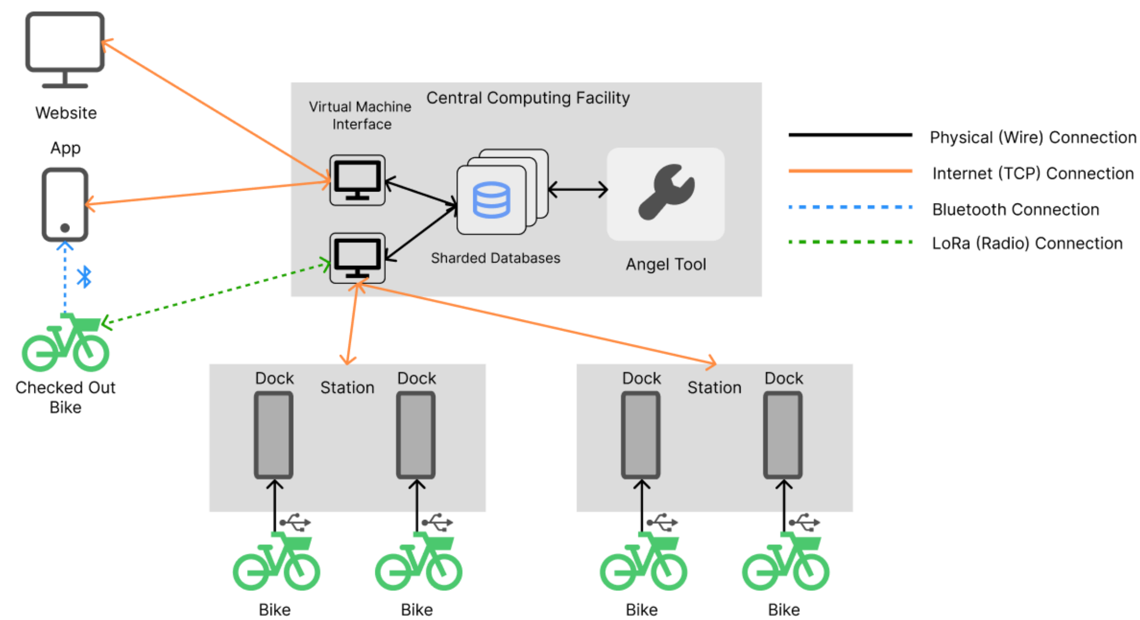
1. Introduction

In urban environments worldwide, the demand for sustainable and efficient transportation solutions has intensified, prompting the development of innovative bike-share systems. However, existing systems often grapple with inefficiencies stemming from disjointed information flow between system modules, resulting in inconvenience for both users and the city. In response to this challenge, we introduce SUPER: Seamless Urban Pedaling: Efficient and Reliable, a pioneering bike share system designed to revolutionize urban mobility.

SUPER addresses the shortcomings of traditional bike share systems by seamlessly integrating information flow across its modules, optimizing user experience and system efficiency. At the heart of SUPER's design philosophy are two primary principles: availability and reliability. Availability, in the context of SUPER, signifies the system's commitment to ensuring that accessing a bike requires minimal effort for users at any given time. Meanwhile, reliability entails the system's ability to consistently deliver a dependable biking experience to users, free from unexpected failures or disruptions.

This paper presents a comprehensive exploration of SUPER, organized into several key sections. First, we provide an overview of SUPER's overarching architecture and operational framework. Subsequently, we delve into the details of SUPER's modules, demonstrating how each component contributes to realizing the system's overarching goals of availability and reliability. Following this, we examine various use cases to illustrate the real-world applicability and impact of SUPER in diverse urban contexts. Finally, we conclude with a discussion on the future implications of SUPER and potential avenues for further development and refinement.

2. System Overview



SUPER is organized around the Central Computing Facility, the machine at the heart of the system. SUPER uses various communication methods to connect the different components, with the user interfaces connecting to the physical devices through the Central Computing Facility. There are 5 main components in the system:

- Central Computing Facility (CCF): All the system's data is stored and aggregated here. This is the center of communications of the system and interacts with all the other components through a virtual machine interface, while periodically communicating with the angel tool.
- Bikes: These are the mobile parts of the system that generate data during rides and communicate with the system through the application and the docks at stations.
- Stations: This is where most of the ride data goes through to arrive at the Central Computing Facility and reservations are made/managed.
- Application: The bulk of user interaction with the system happens here and ride data is accessible to users. The primary communication link is with the Central Computing Facility.
- Website: This only communicates with the Central Computing Facility to allow for limited user interaction with the system (compared to the application).

3. System Design

3.1 Bike

Bikes are the main product to consumers from our business, so it is necessary to ensure they are easy to use, accessible, and reliable. These bikes can be found at stations and there are three types with different electronic capabilities. Users can rent these bikes at stations and by using their mobile or web application.

3.1.1 Data Store and Transfer

Every bike will store minimum information including bike health and GPS data, as well as videos for e-bikes and standard bikes. All data will be transferred from the bikes to the station whenever the bike is docked, resulting in some waiting time between the dock and the next rentals for bikes with videos. However, this mitigates the risk of data loss, prevents interruptions of file transfers, and removes storage limitations on the bikes.

To enable live location data of the e-bikes and standard bikes, we opted for Bluetooth as the wireless communication option, sending GPS readings every 5 seconds to the user's phone. The phone will upload GPS readings, including a timestamp and user ID, to the CCF using TCP. Since GPS data are small in size, we prioritized the accuracy and time-sensitivity of location data over reducing network congestion.

Considering the bike's max speed and predictability of movement, the bike will only save 1 GPS reading per second, about 1/10th of max frequency. This allows bikes to store GPS data for more than 2 hours, which is the longest trip time allowed for users. If a bike is undocked for longer than 2 hours, the LoRa radio will turn on and begin sending the current GPS location, timestamp, bike ID, and user ID every 30 minutes for rescue.

3.1.2 Critical Health

When a bike is damaged users will be notified by the two LED indicators which will blink rapidly. This will prompt users to return a bike to the nearest station to rent a new one. For members with mobile apps riding standard or e-bike, they will receive a notification on their phone directing them to the nearest station via Bluetooth. However, for basic bikes, users will have the option to report a failure through the app and will receive directions to the nearest station to obtain a replacement bike.

3.2 Stations

The stations serve as the only access points to bikes, so it is crucial to make sure they are reliable and accessible, and last for extended periods even under system failures. At stations, all users will be able to use the kiosk to access bikes and see the status of nearby stations, such as the number of bikes and available docks to help users decide as to where their destination will be.

3.2.2 Storage

Each station will locally store information about reserved bikes and docks, as well as dock and undock with bikeID and userID during outages, which will refresh every 5 minutes or as soon as outages are resolved to keep accuracy and consistency. Data related to the bikes such as docking, and undocking times, videos, and GPS information are transferred to the CCF to ensure eventual consistency within the system and clear local data to keep storage available.

3.2.3 System Failures

During power failures, stations will continue operating as they normally do except for the station charging capabilities for Standard and E-Bikes to remain operational for longer than 2 hours.

During communications failures, stations will operate with locally cached reservation data. Even though members cannot make new reservations with limited communication, existing reservations will be seamlessly handled without disrupting the user experience. Any other transactions that do not require data from CCF such as non-member rental will continue to work locally without synchronizing with CCF until the failure ends. In the case local storage is full, the bikes will not be available to rent until the outage concludes and enough space is freed in the local storage.

3.3 Central Computing Facility

The central computing facility is the main source of information for the system and is the central location through which different components of the system communicate. Additionally, CCF is responsible for storing data about riders and trips to manage bike rescue and the angel tool.

3.3.1 Storage

The data will be stored in an SQL database for efficiency and scalability, facilitating data aggregation and analysis for angel tool and traffic data for the city. For security, CCF will store hashed passwords and encrypted personal information encryption with a private key.

Considering the same members will be mostly using the same stations, we will use database sharding to split and store the data in different partitions by station. This improves scalability with smaller partitions to query the appropriate data. The system also uses a centralized controller in software to optimize traffic to and from the different shards based on the station ID in the request.

Additionally, software caches will allow for quick access to regular users’ data in a station. There will be a cache for each station partition. To enforce the reliability of the system, the memory will be written to the disk every 3 hours so that important data is not lost during a failure in the system, but the syncs will not interfere with user interaction with the system.

3.3.2 Communication with Angel Tool

The CCF is responsible for providing riding trends and data to the angel tool so it can determine which bikes need to be moved to which stations. In order to maximize the availability of bikes at all stations, the system will query data regarding availability every hour for the angel tool to update its model to meet the current supply and demand of the system. The data will include the types and number of bikes at each station, the number of users who searched for bikes, the number of reservations at each station, and trip statistics.

Every 15 minutes after making such an update to the Angel tool, the system will receive and implement the suggested bike routes from the Angel tool and add these routes to the mobile application for riders to see. This 15-minute offset is to minimize overall delay for users who may interact with the system during the data aggregation and update period.

3.3.3 Abandoned Bikes

In order to maximize the availability of bikes for all riders, the system focuses on recovering abandoned bikes quickly. The maximum amount of time a rider can rent a bike is 2 hours. Users will also receive a notification near the two-hour mark reminding them to return their bikes. After 2 hours, the bike is considered ‘abandoned’, so the central computing facility locks the bike and receives a GPS location from the bike every 30 minutes using the LoRa radio. If this location remains the same for two consecutive updates, the system will send the bike’s location information through the app for users to rescue the bike. Once a hero rescues a bike and scans the QR code, the CCF will unlock the bike and let the hero ride it to a station.

3.3.4 Expected Datasets

The CCF is expected to store various large datasets to be used by the system to improve its services, the city to analyze its traffic efficiency and users for their own needs. The main categories of these datasets include users, trips, and videos.

The most frequently used datasets in the system will be the user datasets. This includes information about each user such as personal information for authentication, their preferences on stations for availability, and their usage trend for user experiences.

Alongside these details, trip datasets will provide high-level trends of the system for long-range decisions. Information like the number of rides completed, ride time, total amount paid, and number of reservations will be utilized to make system-wide decisions on maintaining hardware or angel tool components and city-wide decisions on traffic control.

Finally, the CCF must store user videos and accident videos, which can be up to 160 GB (45 MB/s * 60 s/min * 30 min * 2 lenses), so this has the potential to grow large quickly. To prevent too much storage use, personal videos will be available for users to download for 24 hours and then removed from the system. Accident videos will be saved on the system automatically and can only be manually deleted if Bikes4All deems it appropriate. These can also be viewed by the user through the application.

3.4 Mobile & Web Application

The mobile and web application serves as the primary interface for users to interact with the bike share system. It encompasses various functionalities aimed at providing a seamless experience for both members and non-members, prioritizing accessibility and ease of use. Members will have full access to both mobile and web applications and non-members will have limited access to web applications.

3.4.1 Functionality

As the most accessible and convenient method of communication, mobile and web applications will be the primary tools for the users to communicate with our system. These are the main functionalities and information that will be communicated:

- Authentication - since it is important to identify users when bikes are stolen or damaged, our system requires personally identifiable information
    - email
    - name
    - password/token
    - address
    - phone number
    - birthday
- Angels & Heroes - this functionality depends heavily on location data including user, bike, and station
    - bike ID
    - user ID
    - user’s current location
    - source location (station or long/lat)
    - destination station
    - time to complete by
    - time requested
- Pickup/Dropoff Reservations - this functionality requires real-time data, so will be synchronized frequently
    - user ID
    - station
    - bike type
    - scheduled time
    - station status such as available bikes and types

- Video Download - users will be allowed to view their personal or accident videos during the trip by downloading them to their mobile devices from CCF within 24 hours of the trip
    - trip ID
- Notification - mobile application will serve as the primary communication between the system and the user
    - title
    - message
    - time notification sent
    - user ID
- Payment - this information will only be communicated once and the following transactions will reuse established payment authorization and tokens, instead of credit card information, for security
    - credit card information

3.4.2 Storage

Given the constraints of limited storage and computation resources, the mobile and web application prioritizes the storage of essential information necessary for a smooth user experience. These include authentication sessions for verifying user's access and caching mechanisms to store frequently accessed data, such as user profiles and recent trip history, enhancing responsiveness and reducing network latency. Cached data are periodically refreshed every hour to maintain accuracy and consistency.

For more details about the authentication session, it will utilize the OAuth2.0 framework of token-based authentication following the JWT standard, which requires two tokens, an access token and a refresh token. Access token is granted when a user authenticates with a short expiration of 60 minutes to avoid abusing credentials or cached tokens by malicious actors and it is attached to every HTTPS request to the server to verify the user. Refresh token, which will last about 30 days, will be used to refresh expired access tokens since it is less exposed to the network. This modularity of authentication gives more control and security to the overall system.

3.4.3 Communication

Given the sensitivity of the data exchanged through mobile and web applications, a strategic approach is adopted to minimize points of failure and maintain data consistency across the system. Communication of the applications is streamlined with the central computing facility using TCP and standards/e-bikes using Bluetooth.

The primary data transmitted, including trip data and user profiles, consists of small-sized text payloads that are not time-sensitive. To uphold the confidentiality and integrity of these data exchanges, a robust encryption mechanism is implemented. Specifically, data is encrypted and verified in JSON format, enhancing security measures. Furthermore, secure communication protocols are enforced, utilizing HTTPS over TCP to safeguard data during transit. This ensures that sensitive information remains protected throughout the communication process, mitigating the risk of unauthorized access or tampering.

3.1. Reliability

SUPER ensures the reliability of the system by ensuring the data across modules are consistent and the user can consistently depend on the functionality of the system, even in moments of temporary failure. To achieve this, the CCF manages all communication and storage in the system with minimum distributed data for efficiency and operation during failure.

SUPER's use of database sharding also ensures the system can scale reliably without putting too much strain on the CCF, so users will have a smooth experience even as the system grows.

3.2 Availability

In our bike rental system, ensuring optimal availability involves a multifaceted approach geared towards seamless access for users. By storing data such as usage trends over time and the history of availability by stations, this system caters to user preferences and demands, ensuring a steady supply of bikes at all times with proactive load balancing and resource allocation.

To further bolster availability, the central computing facility will use real-time GPS data from bikes and members to facilitate rescue and load balancing with corresponding rewards. For example, when there is a big event, such as a Red Sox game, where many people are traveling to the same station at once, we will activate the pre-trained angel tool as soon as we detect the real-time group pattern to load balance as people are arriving.

4. Use Cases

SUPER's reliable and accessible implementation of the bike-share system provides seamless experiences for several use cases:

- Commuters: Sharding will allow for quicker access to a user's information in areas they most frequent. This will improve reliability and will help commuters have more seamless trips.
- External Research: By using frequency and their deviation to most efficient routes we can classify routes as for commute, errands or recreation. Most frequent and efficient will be labeled as commuting, less frequent and less efficient will be labeled as recreational, and those in between as errands. These can be used by the government and other external researchers to better improve road management and recreational facilities.

- Tourists and recreational riders: Through the recording of recreational rides, we have the opportunity to introduce features that enhance the overall recreational riding experience. These will include an option for members and non-members to start a recreational ride, which will have a premade route based on average recreational rides between both stations.
- Underserved communities: In order to maintain fairness and uniformity throughout all stations, information such as the number of bikes, their types, and times of congestion will be used by the Angel algorithm. Using bike type as one of the parameters will ensure that everyone has fair access to all types of bikes.

5. Conclusion

The SUPER bike share system aims to give its users a reliable and accessible bike share system. The improvements we made to our availability and the data we send to the angel system, will ensure availability in peak hours and allow for all users to be treated fairly by the system. Despite communications and power outages, our enhanced protocols ensure that users can continue enjoying the convenience of our bikeshare system for extended durations during these disruptions. The use of database sharding and splitting into partitions allows for increased speeds when fetching user data and can make the experience for those who rely on the system the most even more convenient. Features that can enhance the user experience are also planned, such as giving users using Bluetooth the ability to show up on a map with personalized icons (similar to Waze), and a friend/party system for group rides. SUPER's robust protocols, including enhanced availability measures and innovative features, ensure a reliable and accessible bike-share system, allowing users to seamlessly navigate Newplace even during disruptions.