Solution Idea Title: Smart passenger tracking system for public transport
Planned Pilot Project Duration: 36 months
Main Contact:
Name: Prof. Tara Ghasempouri, Computer Systems Department
Email: tara.ghasempouri@taltech.ee
Phone: +37254637278
University: Tallinn University of Technology

1. Urban Challenge and Problem

Our goal is to improve the efficiency and environmental impact of public transportation in crowded cities by optimizing bus routes via better data collection and analysis. Currently, in most of these cities the collected data can only determine when and where passengers board buses (check-ins) but lack information on where they disembark (check-outs). Moreover, since many of these cities offer free public transportation, many passengers do not have to validate their tickets. Therefore, this missing data makes it difficult to adjust public transportation routes and schedules to meet real demand, reducing system efficiency and contributing to unnecessary CO_2 emissions. In addition, a data analysis method that is effective, real-time, and straightforward to implement is necessary to enhance public transportation. Furthermore, the challenge requires addressing data privacy regulations while collecting and analyzing this passenger flow information. Any solution must ensure that personal data is protected while still providing the city with useful insights.

At the moment the pilot cities of the project are Tartu, Estonia; the main contact is **Jaanus Tamm**, project manager at **Tartu City Government** and Jūrmala, Latvia; the main contact is **Arta Macijevska**, **head of mobility unit Jurmala**. Both of the cities have already determined the above challenges as their issues in public transportation and are eager to find an effective solution.

2. Overview of the proposed solution

To address the issue of incomplete data collection and analysis on the use of public transportation, we propose the implementation of a smart passenger tracking system. The current system in Jūrmala and Tartu only collects data on where passengers board buses, but it does not provide data on the places at which passengers disembark. Due to this limitation, a data analysis method cannot be implemented, and as a result, public transportation routes cannot be effectively optimized.
Our method is composed of three main steps: Data collection, Data storing and Data Analysis.

2.1. Data Collection:

Our solution utilizes **smart cameras** with **edge computing** technology to enable real-time, privacy-compliant data collection:

- **Smart Cameras:** These cameras, installed on buses, are equipped with imaging sensors and GPS modules. They can count the number of passengers boarding and disembarking at each stop, without recording any personal information or images. The image ensures passenger privacy while still accurately tracking movement.
- **Edge Computing:** Data processing happens directly on the bus through edge computing, which reduces the need for transmitting large amounts of data to a central server. This allows real-time data processing and immediate reporting of where and when passengers get on and off the buses.
- **GPS Integration:** The cameras are equipped with GPS systems that accurately track the bus's location. This ensures that not only are the passenger numbers tracked, but the exact location of boarding and disembarking is also recorded.

2.2. Data Storing:

A secure data storage method is implemented in order to ensure the privacy of passengers and to comply with data protection laws:

- As a data storage solution, we recommend using free, scalable database solutions such as PostgreSQL or MySQL. Open-source platforms are cost-effective, widely supported, and can be replicated easily in other cities. We offer a solution that does not rely on proprietary systems, allowing municipalities to maintain full control over their data and to easily integrate it with their existing infrastructure. Cloud-based storage ensures secure data handling, with only the necessary metadata being preserved for analysis.

2.3. Data Analysis:

The collected data is then analyzed using advanced techniques to provide actionable insights for optimizing bus routes:

- **Association Rule Mining Analysis**: In this solution, association rule mining (ARM) is used to uncover patterns and relationships between key factors such as the number of passengers, the time of day, day or night, the location of specific bus stops, and the length of the bus. *As an example, ARM may reveal that if a 12-meter bus carries more than 50 passengers between 7 AM and 8 AM, the majority will disembark at Stop C.* This insight can be used to optimize transportation efficiency. Using a dataset with columns such as Number of passengers, Time (split into 24 intervals, each representing one hour), Day or Night, and Bus Station Stop, information can be gathered to identify high-demand routes and make adjustments accordingly. Our insights allow us to optimize resource allocation by deploying longer buses during peak hours and smaller ones during off-peak hours, thereby reducing operating costs. ARM helps guide bus frequency adjustments, which prevent overcrowding and minimize passenger waiting times at high-demand stops. Additionally, ARM enhances traffic and congestion management by identifying peak disembarkation times and locations, allowing buses to be rerouted to avoid congested areas, improving overall traffic flow and reducing delays. Even holiday or major event spikes can be detected by ARM, allowing city planners to allocate resources more efficiently during these periods. Our solution allows city planners to dynamically adjust bus routes, vehicle sizes, and schedules based on real-time and historical demand patterns. In addition to improving operational efficiency, this approach enhances the passenger experience and reduces overall expenses. It also facilitates the optimization of bus schedules by deploying more buses between busy stops during peak periods, thereby improving the service to passengers. Our recommendation is to use free, open-source database solutions like PostgreSQL that are reliable, cost-effective, and easy to implement and maintain across other cities. We recommend using low-cost tools such as Grafana or Metabase for the visualization of data. As well, since these platforms integrate seamlessly with various database systems, they can connect to existing municipal platforms, eliminating the need for new software and ensuring seamless adoption.

- **Machine Learning Algorithms:** In addition to Association Rule Mining (ARM), we will implement machine learning algorithms leveraging historical data in order to predict where passengers are likely to disembark based on their boarding patterns. A machine learning model can detect patterns and trends in previous trip data, such as passenger numbers, times, and locations, to predict future passenger behavior with high accuracy. It can identify regular travel patterns, such as passengers who board at Stop A during morning rush hour and disembark at Stop C. Over time, the system will learn these patterns and predict future disembarkation points. As a result of this predictive capability, the system is able to adjust bus schedules dynamically to match actual passenger demand. By predicting where and when buses will be most needed, the system is able to allocate resources more efficiently. For example, during peak periods, the system may predict higher passenger volumes on certain routes and recommend increasing bus frequencies. Conversely, during off-peak hours, when fewer passengers are expected, bus frequency can be reduced to save resources without compromising service. Furthermore, machine learning can help optimize routes by identifying underutilized routes that can be consolidated or rerouted, reducing the number of empty or low-occupancy buses on the road. The purpose of this is to improve operational efficiency and reduce waiting times for passengers by deploying buses where they are most needed. Therefore, a public transportation system is adaptive, efficient, and responsive to real-time and historical demand.

- **Real-time Data Integration for Decision-Making:** By continuously collecting and processing real-time data, our system provides city planners and transportation authorities with up-to-date information about public transportation. Metrics such as passenger counts, bus occupancy, and route efficiency are included in this category. In order to visualize this data, we recommend using free or low-cost tools such as Grafana or Metabase, which can be integrated with existing municipal systems without requiring the installation of new software. By using real-time data, planners can make immediate decisions, such as adjusting bus frequencies based on demand or rerouting buses to avoid congested areas. As an example, if certain routes are overcrowded, more buses can be dispatched, or frequencies can be reduced during off-peak hours. By dynamically adjusting to live traffic and passenger flow, the system can also help prevent overcrowding and improve service. With time, the data will aid in strategic planning, optimizing routes and schedules to better meet the needs of the city, resulting in greater efficiency and improved passenger satisfaction.

3. Key Benefits and Innovation of the Solution:

Enhanced Data Collection: The system provides a complete picture of passenger movements, addressing the current gap in data on where passengers disembark. This helps optimize bus routes, reduce congestion, and improve service reliability.

Privacy Compliance: By using image and storing only anonymized metadata, the solution fully complies with privacy laws, ensuring that no personal information is recorded.

Real-Time Optimization: With edge computing and GPS integration, city planners can access real-time data on passenger flows and adjust bus schedules dynamically, reducing idle times and improving efficiency.

Scalability: By using open-source database solutions and flexible visualization tools, the system is designed to be scalable and easily integrated with existing platforms, making it adaptable for other cities and transportation systems.

Justification for the Importance of data mining and machine learning: This analysis is essential for improving the efficiency of public transport in the mentioned cities. By identifying patterns in passenger behavior, the city can tailor bus routes and schedules to meet actual demand. For example, if it is known that certain buses have a high likelihood of being underutilized at specific times, routes can be adjusted to reduce unnecessary trips, saving resources and minimizing environmental impact. Similarly, routes with high demand can be given additional support to prevent overcrowding. The ability to make data-driven decisions through ARM is critical for ensuring that public transportation serves citizens effectively and reduces CO_2 emissions by avoiding underused buses.

Piloting Process

Tartu, Estonia: Tartu's public transportation system includes **67 buses** in **12-meter** and **18-meter** lengths. Our goal is to optimize routes and reduce congestion by using real-time data on passenger movements. A key focus is cutting down **connection times**, ensuring smoother transfers and less waiting for passengers. By adjusting schedules and routes based on actual demand, we aim to make travel faster, more reliable, and convenient for everyone.

Jurmala, Latvia: Jurmala faces challenges with tracking where passengers alight from buses. Implementing our system will allow for better data collection and data analysis to improve route planning and bus schedules while maintaining compliance with local data privacy regulations.

Research Team Capabilities

This project will be developed in collaboration with TalTech researchers, particularly those with expertise in data science, mobility, and privacy regulation. **Prof. Tara Ghasempouri** has strong experience in data mining, applying this science to a wide range of fields, including self-driving buses, testing and verifying digital systems, and securing hardware components. As a machine learning and data mining researcher (PhD holder), **Dr. Mahtab Shahin** has been known for advancing the approaches in machine learning and machine learning through clustering and cloud distribution. Moreover, **Dr. Karl Janson** and **Dr. Uljana Reinsalu**, with background in embedded systems reliability and testing of AI-based embedded systems such as developing smart elevator equipped with cameras. *Karl and Uljana will carry out step 2.1 and step 2.2 of the proposal, Tara and Mahtab respectively, carry out step 2.2 and 2.3.*

4. Expected Impact

Impact on the Environment

By optimizing bus routes and improving efficiency, this system will significantly reduce CO2 emissions by cutting down the number of underutilized buses on the road. The system also enhances the overall quality of public transportation, making it a more attractive option for residents. This encourages a shift towards greener transportation modes, such as walking, cycling, and increased use of public transit. These changes support Jūrmala's and Tartu's long-term goals of becoming more sustainable and environmentally-friendly cities.

Impact on Citizens

Enhanced bus services, including better scheduling and shorter wait times, will lead to higher customer satisfaction. With more reliable public transportation, citizens are more likely to choose buses over private cars, reducing traffic congestion and lowering pollution levels in urban areas. This shift improves both the quality of life for residents and the overall efficiency of the city's transportation network.

Impact on Governance

The system equips Jūrmala and Tartu with real-time data, enabling city officials to make informed, data-driven decisions about transportation management. This will lead to improved operational efficiency and more responsive public services. Additionally, the successful implementation of this solution will serve as a model for other cities looking to adopt smart transportation technologies to enhance mobility and sustainability in urban environments.