

ABSTRACT

Real-time bus occupancy estimation is essential for improving public transit system efficiency and passenger experience. This research investigates the potential of using Wi-Fi probe data for estimating bus occupancy despite challenges encountered due to the randomization of MAC addresses in modern devices. Wi-Fi was identified as a cost-effective and scalable solution after evaluating various technologies, including Automatic Passenger Counter (APC), Automatic Fare Collection (AFC) systems, and crowdsourcing. The study proposes a data-driven approach utilizing the XGBoost regression model to predict bus occupancy. A pilot study was conducted on the Evergreen bus route in Tallahassee, Florida, and the predictions were validated using manually collected occupancy data as ground truth. Two sets of features were used for the model: the first group included all Wi-Fi frame features, while the second group focused specifically on features linked to mobile devices. The model's performance was evaluated across different time intervals. The model yielded an R^2 value of 0.84. The results demonstrate the feasibility of using Wi-Fi probe data for real-time bus occupancy estimation, providing transit agencies with a low-cost alternative with minimal hardware requirements. Future improvements in filtering non-passenger data and noise reduction could further enhance the model's accuracy.