Paper ID #44762

Inferring the Relative Location of Assets utilizing Received Signal Strength Indicator Value of Existing Network Architecture

Mr. Daniél Santos Garza, University of the Incarnate Word
Dr. Okan Caglayan, University of the Incarnate Word

Okan Caglayan is an associate professor in the Department of Engineering at the University of the Incarnate Word (UIW). He received his Ph.D. degree in Electrical Engineering from the University of Texas at San Antonio. The scope of his research ranges fr

Michael Antonio Garcia, University of the Incarnate Word

Inferring the Relative Location of Assets utilizing Received Signal Strength Indicator Value of Existing Network Architecture

Michael Antonio Garcia
School of Math, Science, and Engineering
University of the Incarnate Word

Daniél Santos Garza
School of Math, Science, and Engineering
University of the Incarnate Word

Gonzalo De La Torre Parra, Ph.D.
School of Math, Science, and Engineering
University of the Incarnate Word

Okan Çağlayan, Ph.D.
School of Math, Science, and Engineering
University of the Incarnate Word

Abstract

This paper presents a senior capstone design project to design a remote asset tracking and monitoring system platform by using an organization's local network as a cost-effective alternative solution to a traditional global positioning system (GPS). The proposed system utilizes an existing local area network (LAN) infrastructure to train a machine learning (ML) model to predict and map the locations of an asset, such as a university shuttle. The proposed system was developed and implemented by using a Raspberry Pi single-board computer with an external Wi-Fi antenna to collect a pool of media access control (MAC) addresses and the relative signal strength indicator (RSSI) values as a fingerprint for each location. Furthermore, the latitude and longitude (GPS) data were captured at each of the collection points to train the machine learning models. Once the collected MAC, RSSI and logistical data, features were generated, processed, and exported to Elastic. To determine which model was suitable, we overlaid our chosen Decision Trees, Extra Trees, and Random Forest models on a map to visualize any deviations from our initial GPS data points. The results showed that the Decision Trees model performed the best, with most of the predicted points having an acceptable margin of error relative to our collected data. In the field-testing phase, the plan is to attach the prototype design onto a university shuttle to track its routes around the campus. The results will provide the feasibility of the proposed concept, and it will improve our community's transportation needs by providing more efficient shuttle stops on campus. The long-term goal of the proposed collaborative research between Engineering, Computer Information Systems and Cybersecurity students is to provide safe and healthy spaces by integrating

real-time indoor air quality (IAQ) data within the shuttles to support the community in making informed decisions on daily actions such as catching the shuttle timely on campus.

Introduction

The goal of this project is to explore a cost-effective solution for conventional GPS tracking systems through a new approach to asset tracking. The traditional use of GPS for asset tracking requires that an organization pay for installation, and then a monthly fee for maintenance. In addition, the control the organization has over the hardware is dependent on the software provided from the vendor. This project's process requires only one round of training a machine learning model using gathered data points. Once this initial training is completed, the model generates predictions that effectively identify and address any discrepancies that may occur during asset movements, allowing organizations to monitor their assets throughout their journeys and facilitating the implementation of adjustments to enhance operational efficiency. Multiple techniques have been proposed before that make use of CAZAC sequences such as Zadoff-Chu. These techniques are currently constrained to applications that utilize cellular networks[1]. This project's proposed method bypasses the requirement of a cellular network connection by making use of established on-campus network infrastructure. Similar to other proposed methods, this project's approach requires prior knowledge of signal strengths observed from different access points for the defined anchors. Previous works have made use of Cramer-Rao lower bound for RSS-based positioning with a given signal strength value[2]. Additionally, other attempts have been made to locate objects using RSSI values. However, such attempts were indoors and only incorporated the use of RSSI in a controlled environment[3]. Comparatively, this project's approach makes use of tailored machine learning models to estimate the position based on the observed signal strengths for each BSSID. These stochastic models allow for the use of randomization as a feature, not as a bug in comparison to deterministic methods. This allows variability management within the dataset produced by different weather conditions (for each specific anchor). In a campus setting, where assets such as a shuttle would follow a predetermined route, Wi-Fi signals emitted by access points can be used for a dual purpose. Multiple access points emit multiple Wi-Fi signals simultaneously through the same channel, and clients connect to their network of choice by simply activating their Wi-Fi. The same Extended Service Set Identifier (ESSID) is associated with multiple access points, while each Wi-Fi module within the access point holds a unique Basic Service Set Identification (BSSID). We hypothesize that by identifying the signal strength associated with different BSSIDs across select reference points (Anchors) in the campus, a base platform could be established to predict a current location through the observation of different signal strengths. Calculating the distance observed from each of the different Anchors previously associated with a specific location (longitude and latitude), a module could be implemented that predicts a new set of coordinates based on the observed signal strengths from each of the observed BSSIDs.

System Design

In the initial phase of this project, the objective was to collect data from multiple access points around the university's campus and simulate the route of the campus shuttle by driving along it in a car. This allowed live code adjustments to accurately extract the data for the machine learning model. In addition, the Raspberry Pi was able to collect live RSSI values from each of the access points, as visualized in figure 1. While the hardware was successful in collecting the data, it had computational limitations that made it difficult to utilize in processing our data.

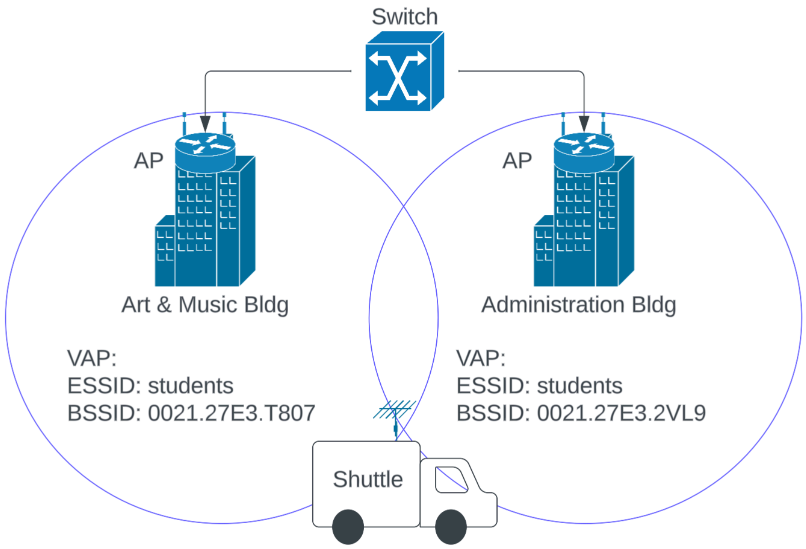

Figure 1 - Diagram showing concept for network-based tracking with campus shuttle

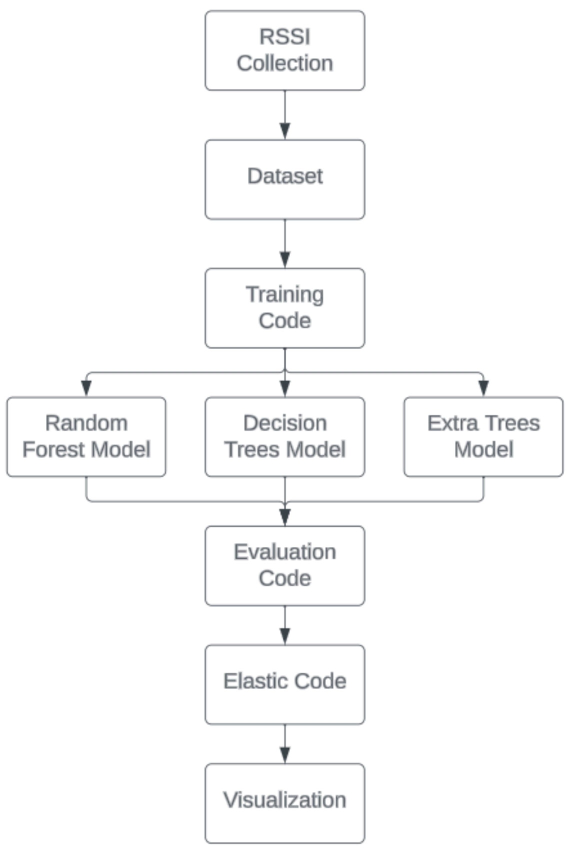

Figure 2 - Flowchart symbolizing the flow of data from collection to visualization

Figure 2 shows the flow of data from the collection to visualization. In order to address the issue of computational resources, a Google Collab environment was utilized to test different machine learning models and training methods. The three training methods that were targeted were decision trees, random forest, and extra trees. In the proposed system, there were a total of six models generated based on the training data of separate latitude and longitude models for each of the training methods. The evaluation code provides readouts of the actual and predicted coordinates in lists. Finally, the Elastic code takes the predicted lists, pairs them together in a document called "predicted_location", then creates an index in an Elastic environment. In the deployment, visualizations were created for each model, with varying degrees of accuracy. It was determined that the Decision Trees training method is the most applicable for our use case.

Methods & Prototype

In this section, we first describe RSSI techniques in detail and their working principles in wireless localization. RSS is the calculation of real signal power received by a receiver, which is typically expressed in decibel milliwatts (dBm) or milliwatts (mW)[4,5]. RSS can be used to measure the distance between transmitter (Tx) and receiver (Rx) devices based on the transmitted and received signal power differences. Generally, two propagation models have been used in RSSI-based wireless sensor networks: (1) free-space models and (2) log-normal models. Free-space models are simple (ideal) but are often limited in real applications because they do not consider obstacles between receivers and transmitters. Therefore, log-normal models are more practical than propagation models and are suitable for indoor and outdoor environments based on their flexibility in different environmental settings[4,6]. Mathematically, the free-space propagation model is defined as follows[4,7]:

$$P_r = 10 \log \left(\frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \right) \tag{1}$$

where P_r is the received power, P_t is the transmitted power, G_t is the transmitter antenna gain, G_r is the receiver antenna gain, λ is the wavelength of the radio waves, d is the distance between the transmitter and receiver, and L is the propagation loss in the channel, which is a function of fading. The log-normal propagation model is defined as follows[4,6]:

$$P_r = P_t + G_t + G_r - L_p(d_0) - 10 \log \left(\frac{d}{d_0} \right)^{\alpha} + C_a. \tag{2}$$

In (2), α is the path loss exponent, which depends on a specific propagation environment, $L_p\,(d_0)$ is the path loss at a reference distance d_0 , and C_a is a normally distributed random number with zero mean and a variance of σ^2 considering shadowing and other sources of uncertainty ($C_a \sim N\,(0, \sigma^2\,)$) [unit: dB][4,6]. The RSSI representing the RSS level at a receiver device has an arbitrary range of values that the chip supplier primarily characterizes. For example, a receiver device may translate dBm values into RSSI values ranging from 0 to 60, 0 to 100, or −100 to 0, depending on the chip vendor. RSSI is one of the simplest and most widely used indoor positioning tools in the literature[4,8], but can be applied as a solution for outdoor asset location. Although RSSI-based solutions have advantages such as lower device requirements, better accessibility, and cost-effective system design, they also suffer from numerous problems in indoor and outdoor environments[4,9]. These problems include significant path loss, multipath fading loss, indoor noise and interference, absorption loss, and the unavailability of some APs during localization. Various building materials also affect RSS levels, as shown in Table 1. To address these issues, several solutions have been proposed in the literature, including various filtering and averaging methods, RSS cutoff and self-calibration techniques, the use of an increased number of APs or reference points (RPs), and ML-based schemes. In particular, ML-based schemes such as wireless signal recognition using ML and channel modeling using ML are promising candidates for solving RSSI-based outdoor and indoor localization issues[4,10-12].

Table 1 - Decline of RSS power level proportionate to material that signal travels through to reach antenna[4,13]

Material	**Decline in RSS power level (dB)**
Plasterboard	3-5
Glass wall & metal frame	6
Metal door	6-10
Window	3
Concrete wall	6-15
Block wall	4-6

The data collection hardware consisted of a single-board Raspberry Pi 4B computer, an LCD1602 I2C display unit, and an Alfa AWUS036H 2000mW Long-Range WiFi USB adapter, as seen in figure 3 below. For the purposes of collecting geographical points, information was consolidated from phones with GPS and a Garmin GPS watch, displayed on a map in figure 4. During early testing it was determined that a more compact solution, such as a Raspberry Pi Zero did not have the computing power necessary to execute the collection code. Additionally, the on-board Wi-Fi module on neither computer was consistently strong enough to detect the signal being produced from Wi-Fi access points. This hardware was able to collect mac addresses, RSSI values over the course of twenty seconds, and GPS coordinates of each successful access point connection site.

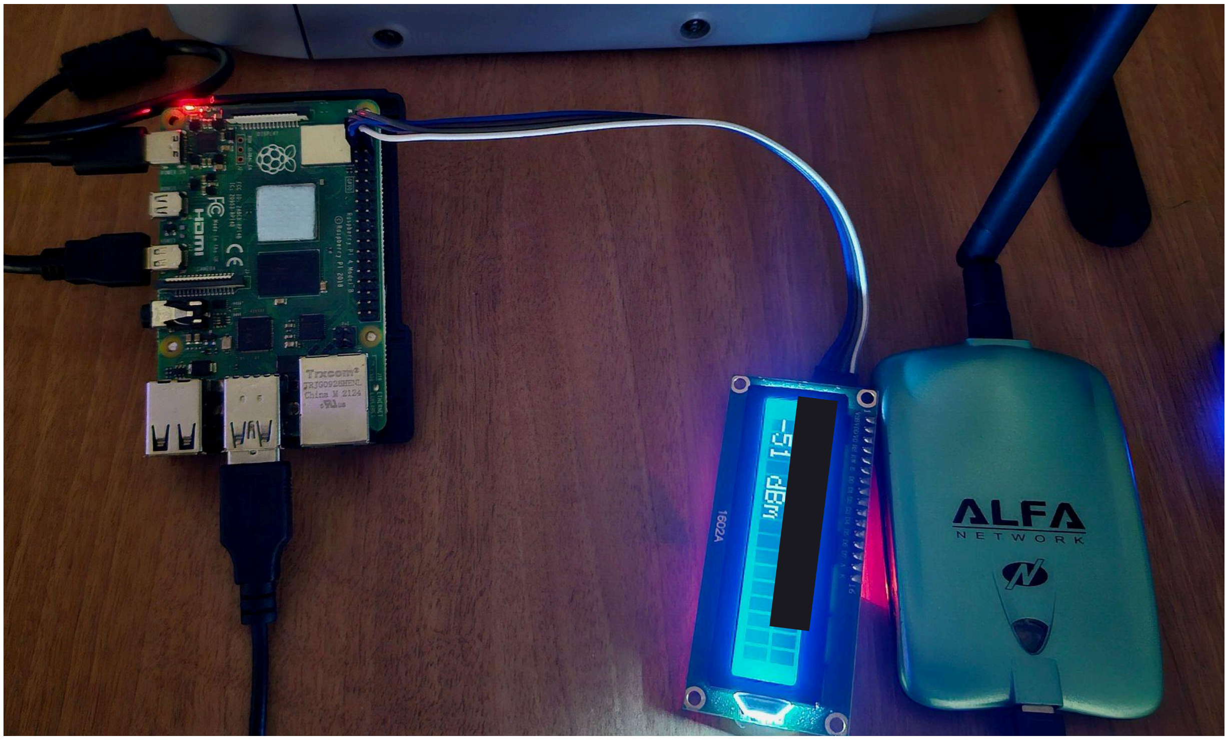

Figure 3 - Hardware for data collection displaying MAC and RSSI values

Figure 4 - Map showing the collected GPS points at every successful access point connection

Dataset
During the data collection process, signal strength information was received from one access point at a time. We used a Garmin watch to determine GPS coordinates, and collected 20 RSSI values from our Raspberry Pi over the course of 20 seconds. The linear relationship between features and numerical values facilitates the models in recognizing this pattern, contributing to predictions with limited ambiguity. Upon examination of the model's predictions against the ground truth, minor differences in the Mean Standard Error (MSE) were noted, as observed in both table 2 and 3. For a thorough comparison with results from other studies, expanding the dataset and retraining the models with additional data is recommended. This approach aims to provide a more realistic MSE and distance error, enhancing the evaluation of our findings.

Table 2 - Mean squared error (MSE) and proportion of variance (R2) for the latitude train (T) and validation (V) stages

Latitude	MSE (T)	MSE (V)	R2 (T)	R2 (V)
DecisionTree	3.946408e-29	3.420501e-29	1.0	1.000000
ExtraTrees	9.174901e-18	9.181227e-11	1.0	0.999809
RandomForest	2.010522e-27	2.010522e-27	1.0	1.000000

Table 3 - Mean squared error (MSE) and proportion of variance (R2) for the longitude train (T) and validation (V) stages

Longitude	MSE (T)	MSE (V)	R2 (T)	R2 (V)
DecisionTree	2.533891e-28	1.913461e-28	1.0	1.000000
ExtraTrees	1.640730e-26	3.026475e-11	1.0	0.999979
RandomForest	1.643708e-26	1.644112e-26	1.0	1.000000

Machine Learning Models
The dataset consists of 500 data points with 25 features, each representing a different level, and 2 labels representing latitude and longitude coordinates. The data were divided into training and testing sets using an 80-20 split, with 80% (400) of the data points allocated for training and 20% (100) for testing. Three training methods, namely DecisionTree, ExtraTrees, and RandomForest regressors, were trained on the preprocessed data. The training process involved 10-fold cross-validation to assess the models' performance on multiple subsets of the training data. Following the completion of the cross-validation loop, we calculate the average R2 values over all folds to provide a comprehensive assessment of model performance. The following figures depict graphic visualizations of our Decision Tree model.

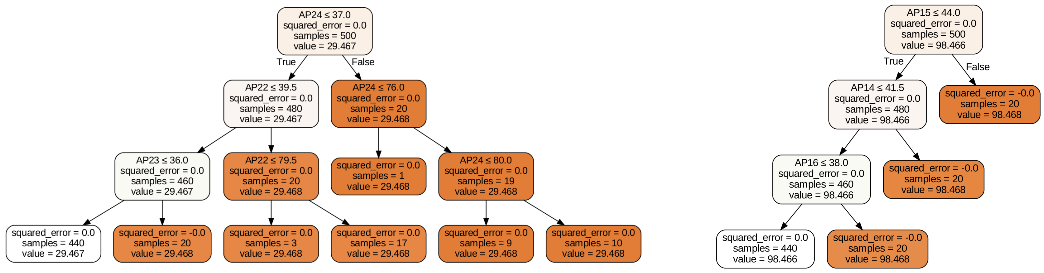

Figure 5 - Visualizations of truncated Decision Tree models. Latitude (left) and Longitude (right).

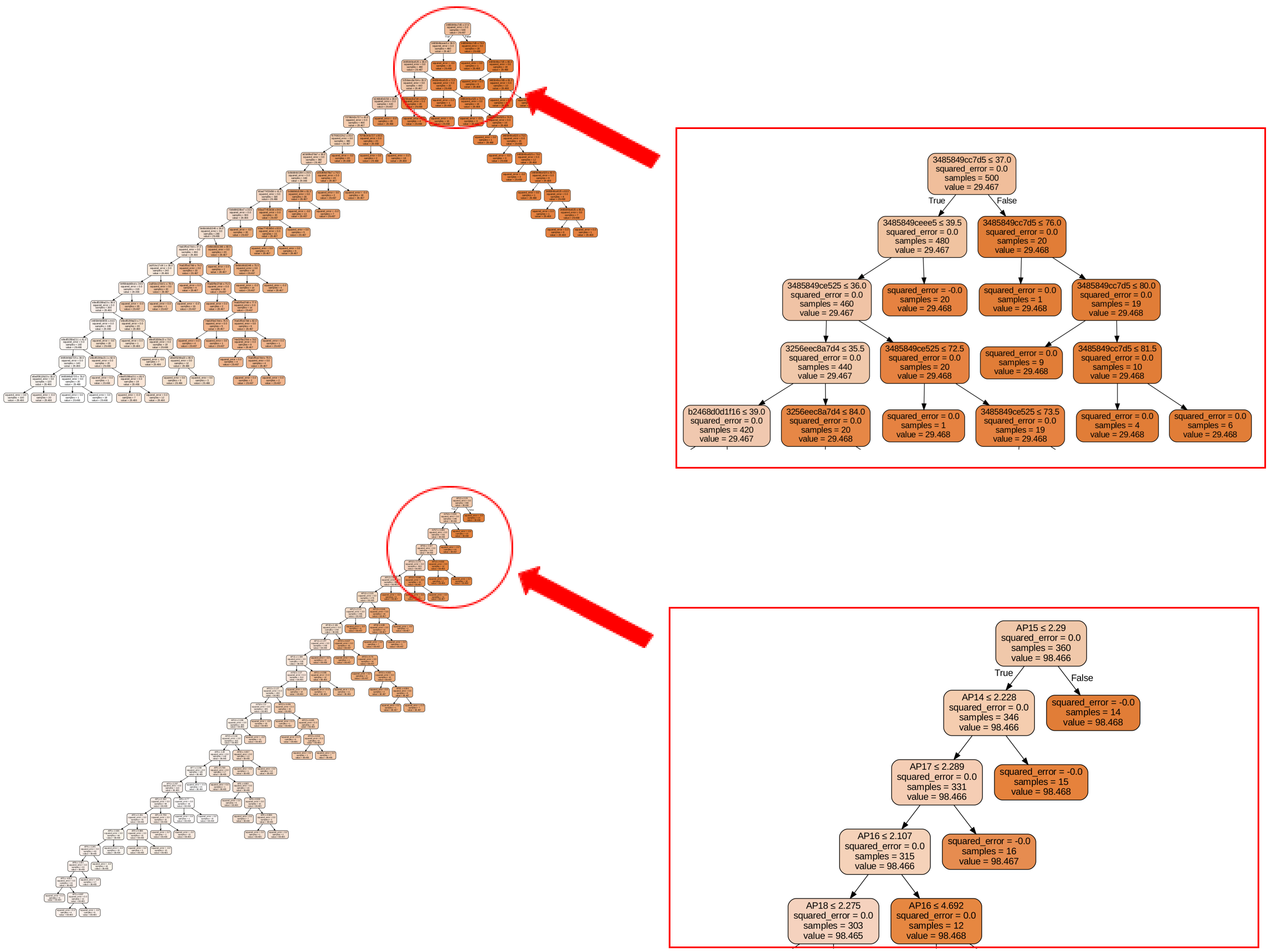

Figure 6 - Visualization of complete models. Latitude (top) and Longitude (bottom).

Results

Each training method generated two models, one for latitude and one for longitude. Using the Elastic python library, both models were combined into the appropriate format, then ingested them into an index created in the code. Despite the relatively simplistic dataset, each model performed with varying degrees of accuracy. As indicated in tables one and two, the decision tree and random forest models performed more accurately than the extra trees model. However, due to the lower MSE value, the decision tree model was chosen. As indicated in figure 7, this model predicted the location with a miniscule difference compared to the original collected coordinate points.

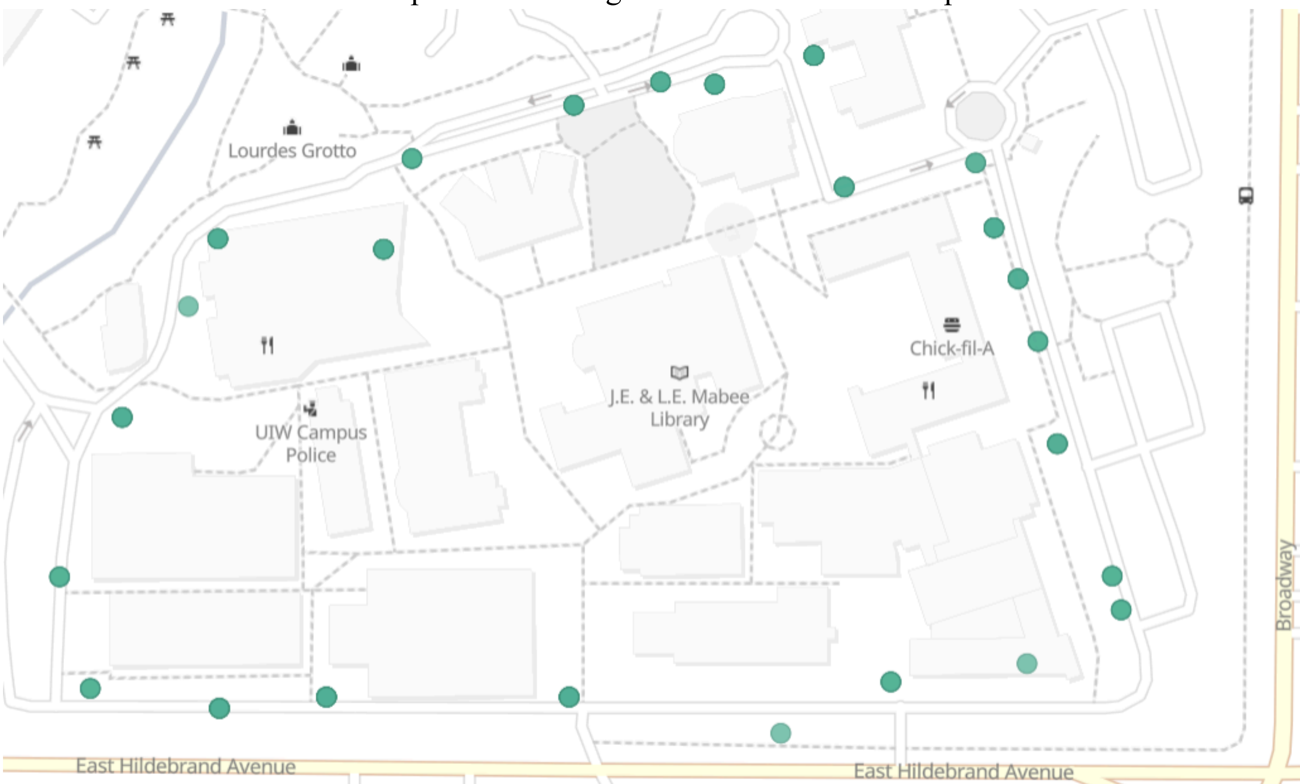

Figure 7 - Map showing the Decision Tree model predicted location at each access point

Conclusion

This project offers a cost-effective solution for asset tracking within an organization's network. It eliminates the need for additional architecture to be installed or a GPS device to be implemented. The initial data collection for the machine learning model sets the baseline, allowing it to generate insights for optimizing asset movement. In its current state, the project is the first phase of a larger system. The second phase of this project involves a more robust training data set as well as multioutput machine learning models. This work will expand into incorporating a previously created Air Quality Index (AQI) device and include its measurements into the generated telemetry[14]. All collected data will then be exported to the cloud for visualization on a computational system. The AQI telemetry generation will be facilitated by an Arduino equipped with sensors, with data exportation integrated into the Raspberry Pi 4B utilized in this project.

References

1. Beyme, S., & Leung, C. (2009). Efficient computation of DFT of Zadoff-Chu sequences. Electronics Letters, 45(9), 461. https://doi.org/10.1049/el.2009.3330
2. Veletić, M., & Šunjevarić, M. (2014). On the Cramer-Rao lower bound for RSS-based positioning in wireless cellular networks. AEU - International Journal of Electronics and Communications, 68(8), 730–736. https://doi.org/10.1016/j.aeue.2014.02.012
3. Li, G., Geng, E., Ye, Z., Xu, Y., Lin, J., & Pang, Y. J. (2018). Indoor positioning algorithm based on the improved RSSI distance model. Sensors, 18(9), 2820. https://doi.org/10.3390/s18092820
4. Machine Learning based indoor localization using Wi-Fi RSSI Fingerprints: an overview. (2021). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/document/9531633
5. RADAR: an in-building RF-based user location and tracking system. (2000). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/832252
6. García, J. V. (2020). Characterization of the Log-Normal model for received signal strength measurements in real wireless sensor networks. Journal of Sensor and Actuator Networks, 9(1), 12. https://doi.org/10.3390/jsan9010012
7. Yamamoto, B. E., Wong, A. R., Agcanas, P. J., Jones, K., Gaspar, D., Andrade, R., & Trimble, A. Z. (2019). Received Signal Strength Indication (RSSI) of 2.4 GHz and 5 GHz Wireless Local Area Network Systems Projected over Land and Sea for Near-Shore Maritime Robot Operations. Journal of Marine Science and Engineering, 7(9), 290. https://doi.org/10.3390/jmse7090290
8. A survey of indoor localization Systems and technologies. (2019, January 1). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/document/8692423
9. A joint indoor WLAN localization and outlier detection scheme using LASSO and Elastic-Net optimization techniques. (2017, August 1). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/7600441
10. Wu, C., Yang, Z., Zhou, Z., Liu, Y., & Liu, M. (2017). Mitigating large errors in WiFi-Based indoor localization for smartphones. IEEE Transactions on Vehicular Technology, 66(7), 6246–6257. https://doi.org/10.1109/tvt.2016.2630713
11. Li, X., Dong, F., Zhou, S., & Guo, W. (2019). A survey on deep learning techniques in wireless signal recognition. Wireless Communications and Mobile Computing, 2019, 1–12. https://doi.org/10.1155/2019/5629572
12. Cisco Unified Wireless Location-Based Services [Design Zone for Mobility]. (2008, September 10). Cisco. https://www.cisco.com/en/US/docs/solutions/Enterprise/Mobility/emob30dg/Locatn.html
13. Implementing 802.11, 802.16 and 802.20 wireless networks. (n.d.). Google Books. https://books.google.com/books/about/Implementing_802_11_802_16_and_802_20_Wi.html?id=qytvAz_VYKQC
14. Chavez, J. E. (2022, March 16). Integrated Multi-Sensor Remote System design for Real-Time indoor air quality monitoring. https://peer.asee.org/integrated-multi-sensor-remote-system-design-for-real-time-indoor-air-quality-monitoring

MICHAEL ANTONIO GARCIA

Michael Antonio Garcia is a student at the University of the Incarnate Word in the School of Math, Science, & Engineering. He is graduating in May of 2024 with a Bachelor's of Science with a double major in Cybersecurity and Computer Information Systems, with a concentration in Computer Programming.

DANIÉL SANTOS GARZA

Daniél Santos Garza is a student at the University of the Incarnate Word in the School of Math, Science, & Engineering. He is graduating in May of 2024 with a Bachelor's of Science with a

double major in Cybersecurity and Computer Information Systems, with a concentration in Computer Programming.

GONZALO DE LA TORRE PARRA

Dr. Gonzalo De La Torre Parra is a distinguished Assistant Professor in the Department of Engineering, Computing, and Cybersecurity at the University of the Incarnate Word. With a Ph.D. in Cybersecurity from the University of Texas at San Antonio, achieved in 2021, Dr. De La Torre Parra has an impressive academic and professional background. During his doctoral studies, he notably served as a research fellow, spearheading pivotal cybersecurity projects. These projects, conducted in collaboration with industry giants such as CPS Energy and CISCO, underscore his ability to bridge academic research with real-world applications. His educational journey began with a Bachelor's degree in Electrical Engineering from Texas A&M University-Kingsville, obtained in 2009, followed by a Master's degree in Electrical Engineering from The University of Texas at San Antonio in 2015. Dr. De La Torre Parra's research is profoundly centered on the integration of artificial intelligence into cybersecurity. His work primarily focuses on enhancing security measures in cloud computing, edge computing, and Internet of Things (IoT) technologies. As an eminent figure in his field, Dr. De La Torre Parra has made substantial contributions to the development of Large Language Models (LLMs) for Code Vulnerability Detection. His work not only advances academic knowledge but also has significant practical implications in the industry. His commitment to academia is further exemplified by his dedication to mentoring aspiring cybersecurity professionals. Dr. De La Torre Parra's guidance and expertise continue to inspire and shape the future leaders in cybersecurity, making him a valuable asset to both the academic and industrial spheres of cybersecurity.

OKAN ÇAĞLAYAN

Okan Çağlayan, Ph.D., is an Associate Professor of Engineering, in the Department of Engineering, Computing, and Cybersecurity (ECCS) at the University of the Incarnate Word (UIW). He received his Ph.D. degree in Electrical Engineering from the University of Texas at San Antonio (UTSA). Prior to joining UIW, Dr. Caglayan worked as an engineering consultant in the Applied Power Division at Southwest Research Institute (SwRI). In addition, he was a lecturer in the Department of Physics and Astronomy at UTSA teaching Engineering Physics with emphasis on electromagnetism, mechanics, and optical science. The scope of his research ranges from developing new techniques in the areas of digital signal/image processing with pattern recognition applications and computer vision to building innovative smart sensor and Internet of Things (IoT) networks, low power embedded systems by using machine learning, such as TinyML, in real-time implementations. His current work on cyber-physical systems has led to the design and implementation of ubiquitous sensing, intelligent sensor networks, smart wearable devices, multi-sensor integrated (sensor fusion) embedded devices, wireless sensor networks (WSN), biomedical sensors, cloud/edge-cloud computing algorithms, sustainable (green) power management techniques, such as off grid solar powered systems. Dr. Çağlayan is a member of Institute of Electrical & Electronics Engineers (IEEE) Signal Processing, Image Processing, Computer Societies, American Society for Engineering Education (ASEE), Eta Kappa Nu (IEEE Honor Society), Omicron Delta Kappa (The Circle and ODK) Honor Societies.