**University of Arkansas – CSCE Department**

**Capstone I – Final Proposal – Fall 2016**

Transit Sensor

**Caleb Welch, Christopher Lail, Jackson Haley, Jeff Johnson, Vitaly Borodin, Zaid Masri**

## Abstract - Zaid Masri

Public transportation offers a convenient, affordable, and sustainable mean of transportation. However, public transportation can also be very inconvenient, as many individuals choose to take this form of transportation due to its affordability. This, in turn, causes a limited amount of open seats. These problems are studied and researched by transit planners who are tasked with forecasting long term demands and traffic patterns.

Real time data collection has become increasingly easy with Wi-Fi and Bluetooth-enabled devices. With this form of collection, our data will allow transit planners to help pedestrians establish better wait time and arrival time estimations.

However, data collection through Bluetooth and Wi-Fi can also be problematic and costly which is why we aim to create an inexpensive wireless system that uses off-the-shelf Bluetooth and wireless hardware to wirelessly detect end-to-end passenger journeys.

## 1.0 Problem - Zaid Masri

Public transport reduces the use of private cars, hence reducing fuel consumption and traffic congestion. At the University of Arkansas, the number of people who are choosing public transport over the use of private cars due to cost of parking, as well as saving time on attempting to find a parking spot on campus, is slowly increasing.

However, many students face two major issues with the transit system: potentially excessive wait time as well as the risk of buses reaching maximum capacity before the students are able to board. Transit buses do offer timetables and public services, such as google maps, that improve student’s wait time estimation, however transit and parking still lack a comprehensive method of collecting ridership data. The transit department primarily collects data by manually tracking the total number of passengers who board the bus. Since the time at which a passenger boards, the stop at which they board, and the stop at which they depart aren’t recorded, this type of data collection can lead to a number of inaccuracies, and statistical analysis of this data is of limited value.

## 2.0 Objective - Zaid Masri

The objective of this project is to develop a system to accurately estimate transit ridership data using mobile devices belonging to the system’s passengers. The development of this data collection system will not only provide transit ridership data to better design our public transit systems, but it can also be used to provide useful real-time transit information to passengers. Doing this would provide accurate arrival time of the next bus and allow passengers to take alternative transport choices instead.

Overall, we will develop a low cost hardware and software sensor system to detect a passenger’s unique MAC address or Bluetooth signal and use data post processing to distill passenger boarding counts by time of day, location, and route.This data will then be used to generate passenger origin-destination matrices, wait times, and other metrics of interest.

## 3.0 Background

### 3.1 Key Concepts - Jeff Johnson

* Wireless and Bluetooth devices in close proximity have uniquely identifying information that can be found using their respective protocol’s discovery functionalities. By placing devices capable of recording this data at points of interest in the transit system, namely a combination of buses and bus stops, it is possible to collect comprehensive ridership data for a statistically meaningful subset of the transit population.
* When combined with data taken from the transit department and the transit system API, statistical analysis on the data collected by our device can be used to calculate a variety of useful transit metrics, such as the median wait duration for passengers at bus stops as well as typical origin-destination matrices for passengers of each bus route.

### 3.2 Related Work - Zaid Masri

The collection of real time transit data is an ongoing issue; researchers constantly work on trying to improve how efficiently and accurately they collect data.

In the past, researchers have conducted surveys on board buses and in stations. Bruce Schaller from Schaller Consulting has published a paper in which he discusses the use of surveys to collect data[3]. After collecting the surveys, his team would carry out the task of checking, cleaning, and tabulating the meaningful data. Transit agencies aided Schaller’s team by urging passengers to assist with the research and fill out the surveys.

The raw data that Schaller was successful in collecting was mostly sufficient for his research. However, there were some problems with the system, such as completeness: Schaller needed a certain percentage of the survey to be completed in order to consider it usable, and not all passengers would fill out the survey to completion. Additionally, not all transit agencies enforced editing procedures or other steps that would improve data accuracy.

Pedestrians would often be given questions such as “What would your wait time be?” and they would give estimates rather than accurate data. Therefore, the data collected from surveys becomes more of a theoretical study and a less of an accurate approximation. This form of data collection also proved to be inefficient as Schaller’s team needed to distribute surveys on board different transits and make sure that every driver is knowledgeable of the procedure they are carrying out.

In addition to the aforementioned faults, the fact that this data cannot provide real-time updates or estimations of ridership data led some researchers to begin utilizing Bluetooth-enabled devices to capture anonymous pedestrian data. This proved to be a much more efficient way of collecting ridership data. Because the Bluetooth method also collects data using substantially fewer person-hours, it is also believed to be less costly. These factors have caused the method to become increasingly popular. A group of researchers from Washington University were able to listen to devices in a specific area by utilizing short range protocol identifiers such as Bluetooth Media Access Control[4]. Many individuals carry a personal wireless device that broadcasts their unique MAC address when the device is in “discoverable” mode. For this study the group of researchers installed Motorola Droid devices on high-volume transit stops.

One of their biggest issue was the abundance of very short wait time which is usually under one minute long. These times are usually not a result of just-in-time arrivals but are outliers that can occur in a few different situations. The most common is the occurrence of Bluetooth devices from passengers already on a bus that is stopping at a stop. This creates the impressions that one or many devices have appeared and left with the bus.

Overall, both forms of research were primarily used for calculating pedestrian wait time and with that utilizing that data to improve transit agencies on-time performance.

Our project will consists a combination of all the studies listed. We do plan on distributing surveys on buses in order to collect raw data as well as capturing phone’s MAC address using Wi-Fi and Bluetooth signals emitted by the smartphones. How we aim to capture the MAC addresses is by using a Raspberry Pis which would be installed inside the transit and around high-volume transit stops. After the collecting the necessary data we aim to filter it down and compare each form of data collection methods we used. This would narrow down the “disposable data” and give us a better approximation of ridership data as well as a more accurate wait time. This data could also be utilized for real-time updates.

## 4.0 Design

### 4.1 Requirements and Design Goals - Chris Lail, Caleb Welch

Device should:

* Estimate ridership on transit vehicle at any point in time
* Note the origin and destination of each unique device on the transit system
* Estimate average duration for which passengers wait at a stop
* Communicate the above data to a central location for statistical analysis
* Use a power/connectivity solution that requires as little manual maintenance as possible
* Be designed such that mass implementation is financially feasible

### 4.2 Detailed Architecture - Chris Lail, Caleb Welch

**Functionality**

The base level device functionality will utilize the following protocols to detect rider’s phones and devices:  
**Wi-Fi**

* Wi-Fi is present in most popular smartphones, laptops, and many other modern portable devices, and is used to connect to the internet wirelessly.
* We would make use of this protocol to count devices, as when Wi-Fi is turned on the device is constantly searching for hotspots to connect to, and when it finds one, it will be noted by our device, and counted as a rider. This would be our main method of counting.

**Bluetooth**

* Bluetooth is present in most popular smartphones, and a few other types of devices such as speakers and other accessories. For security purposes, in many cases you must confirm Bluetooth
* Since Bluetooth must be on and discoverable for it be counted, and there is the potential to be non-phone devices, this metric will be inherently more inaccurate, so it will be secondary.

**Hardware**

Our hardware solution must be as cheap as possible while maintaining full functionality, as well as having the flexibility to adopt multiple scanning methodologies. With that in mind, we settled on the following devices:

**Raspberry Pi**

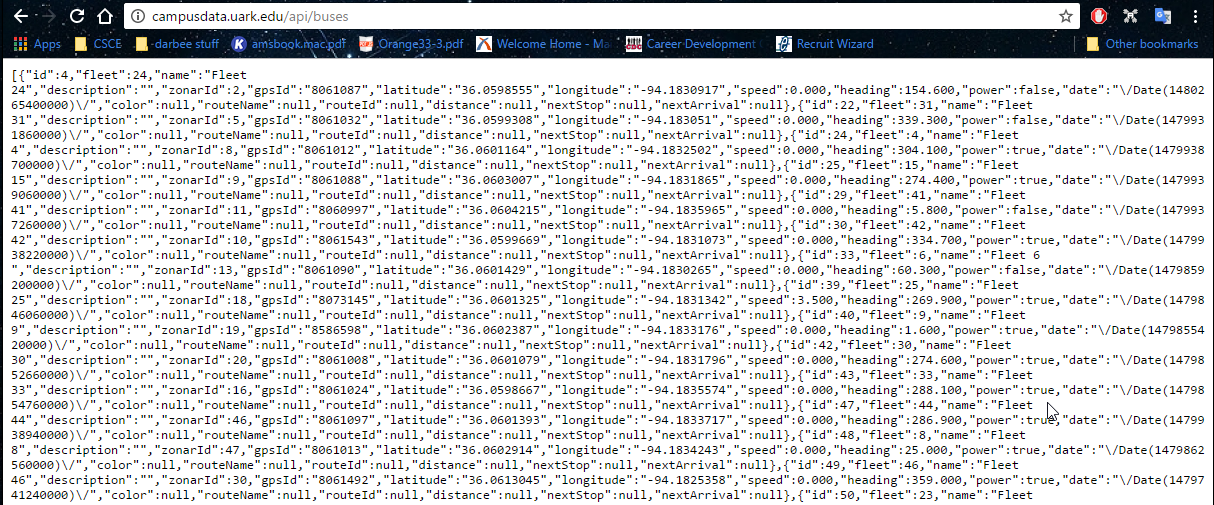
* The Raspberry Pi is a cheap, tiny computer which can run linux or other operating systems and efficiently perform relatively small computations. It has USB, ethernet, SD card, and hardware sensor pin connectivity.
* We are using Raspbian linux on the Raspberry Pi, as it is specifically designed for Raspberry Pi and still can run Python and our specific libraries.
* We are using both a Raspberry Pi 2 and 3 for our prototyping, as well as for our recommended final build we will provide to the Mechanical Department.

**Accessories**

* As the Raspberry Pi is designed to be flexible, it does not come fully featured out of the box and we have to supply it with accessories to make it cover all of our required use cases.
* We have obtained a Wi-Fi and Bluetooth adapter for our older Raspberry Ri 2.
* Lastly, we will need an enclosure for the device. During the initial stages we are using a Pelican water proof hardcase.

**Software**

The software will use a python script to find and store all MAC Addresses from the users connected to the network.

* Using a library called **scapy**, a popular low level packet and network manipulating library for Python, a script can be running on the Pi as long as it is on.
* Using the ***sniff()*** function, the script can detect packets being sent across the Pi’s network. This can grab the Mac Addresses of any users interacting the network and then a timestamp using the system's internal clock can be added to the data stored.
* Using the Pi’s Wi-Fi dongle or built-in Wi-Fi, the Pi will be able to connect to a mySQL server using a library called **MySQLdb**. When the Pi detects a Wi-Fi network it knows, i.e. UARK Wi-Fi, it will secure a connection the our database and upload either a JSON or CSV of the data it has collected.
* Additionally, the UARK Transit System’s API can be utilized to get the location of the busses in order to determine which bus is on which route at any given time, as certain busses service a variety of routes, as well as the busses current speed the next stop which it will service. 
* JSON and CSV files are very easy to manipulate in Python using a library called **Pandas** which can flow into data analysis techniques. Libraries such as **SciKitLearn** and Google’s **Tensorflow** can be used for machine learning techniques. **SciKitLearn** is a very simple and easy to use library although **Tensorflow** is much more robust. After the data is processed, we can use a graphical library like **Seaborn** in order to present the information in a way that is easily readable and more impactful.

### 4.3 Tasks - Jeff Johnson, Jackson Haley, Vitaly Borodin

1. Meet with sponsor to procure provided hardware and materials
2. Begin prototyping transit sensor device
3. Schedule contact with sponsor and transit department for later in the semester
4. Revise device prototype until it is portable
5. Begin designing database server
6. With transit department’s assistance, use portable transit sensor prototype to collect preliminary ridership data
7. Manually collect ridership data and compare that data to transit sensor’s finding, tune device software as necessary
8. Finalize database server and begin data analytics implementation
9. Integrate finalized transit sensor device into transit system with transit department’s assistance
10. Finalize data collection methodology, data analytics, and device documentation
11. Prepare final report and privacy post-mortem

### 4.4 Schedule - Jeff Johnson, Jackson Haley, Vitaly Borodin

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tasks** | **Jan 16 - Feb 17** | **Feb 17 - Mar 18** | **Mar 18 - Apr 19** | **Apr 19 - May 20** |
| Procure Hardware | \_ |  |  |  |
| Prototype Transit Sensor | \_ |  |  |  |
| Contact Transit Department | \_ | \_ |  |  |
| Make Sensor Portable |  | \_ |  |  |
| Implement Database Server |  | \_ | \_ |  |
| Data Collection & Tuning |  |  | \_ |  |
| Implement Data Analytics |  |  | \_ |  |
| Finalize Transit Sensor |  |  | \_ | \_ |
| Documentation | \_ | \_ | \_ | \_ |
| Final Report |  |  |  | \_ |

### 4.5 Deliverables - Vitaly Borodin, Jackson Haley, Jeff Johnson

### Design Document:

* + Documents functionality and details typical usage procedures of our devices.
  + Describes steps necessary to construct and implement our hardware and software so that it can be replicated for additional buses and/or bus stops.
* Mobile Device Scanner
  + Hardware
    - Our physical equipment will consist of a Raspberry PI with Bluetooth and Wi-Fi dongles, potentially with additional hardware for detecting nearby cell phone radio frequency signals.
  + Software
    - Our application will scan within a close vicinity in order to count the number of mobile devices in the area. This data will be used to estimate the actual population of transit riders at that point in time.
    - The scanning will be done with some combination of the Wi-Fi, Bluetooth, and radio frequency scanning equipment associated with our hardware devices.
    - From the string of the Wi-Fi and/or Bluetooth we can calculate the distance to the mobile devices.
    - The software for the Raspberry Pi hardware solution would be Python-based.
    - The software would track the unique identifiers for each Wi-Fi or Bluetooth device along with the locations and timestamps at which they entered or exited the transit system.
    - The software will encrypt the MAC addresses before sending to database server.
    - Additionally, software could possibly utilize radio frequency scanning hardware to accurately report the total number of devices on the bus, rather than just those with enabled Bluetooth or wireless adapters.
* Database for Compiled Ridership Data
  + Table fields will include the scanner’s DeviceID, Bus/Stop ID at which scanner is installed, Passenger device MAC, Passenger Entry Time, Passenger Entry Bus Stop ID, Passenger Exit Time, and Passenger Exit Bus Stop ID.
  + Bus/Stop ID fields can match data from University Transit APIs.
  + Database contents can be passed to web server to perform statistical analysis.
  + The data will be periodically, (ones pre semester) deleted to save space and privacy.
  + API using PHP and JSON to implement database management functionalities.
* Web Server/Interface
  + Make web interface to access data from DB using PHP and JavaScript.
  + Create data processing algorithms to analyze scanning data in order to create passenger origin-destination matrices, determine estimated bus stop wait times, etc.
  + Implement user friendly interface to analyse the data.
* Wireless Scanning Survey Report
  + This document will compare the data gathered using our scanning system with actual transit ridership data gathered by our own research or in combination with the University Transit Department in order to determine how accurately our solution estimates actual ridership data.
* Privacy and Legal Issues
  + Additional report documenting both our research regarding and our experience handling the potential legal and privacy issues associated with this project.
  + MAC addresses are unique identifiers that may be linked to individuals and, consequently, may lead to personal information. To preserve our passenger’s privacy, we will hash all the identifying information we collect before storing it such that those hash values would no longer link to individual.
  + Communication between our devices and server will need to be encrypted and the server itself must comprehensively guard its data.
  + Consequently to prevent any data leakage we will delete all the old data.
  + Finally, a non-intrusive method of disclosure may need to be agreed upon with our sponsor and the transit department in order to inform passengers that their transit information may be being collected.

## 5.0 Key Personnel

**Caleb Welch**– Welch is a senior Computer Science major in the Computer Science and Computer Engineering Department at the University of Arkansas. He has completed Algorithms, Artificial Intelligence, Computer Networks and Software Engineering. Welch had a previous internship doing data analysis on SQL data and used machine learning techniques to benefit the rest of the staff. He will be responsible for the project’s software and post data analysis.

**Christopher Lail** - Lail is a senior Computer Science major in the Computer Science and Computer Engineering at the University of Arkansas. He has completed Algorithms, Computer Networks and Software Engineering, among others. He will be responsible in part for the software and database aspects of the project.

**Jackson Haley** - Haley is a senior Computer Science major in the Computer Science and Computer Engineering Department at the University of Arkansas. He has completed Programming Paradigms, Applied Cryptography, Software Engineering, and Algorithms, among other courses. He has had internships with the University of Arkansas Division of Agriculture Cooperative Extension working on Android applications and with Word Machinery working on a desktop application in Visual Basic and multiple iOS applications, several of which utilize an SQL server. He will be responsible for cooperatively working on both the software for the sensor as well as the SQL database and the web interface.

**Zaid Masri** - Masri is a senior Computer Engineering major in the Computer Science and Computer Engineering Department at the University of Arkansas. He has completed Database Management Systems, Programming Paradigms, Embedded Systems, and Software Engineering. Masri has had research experience with GPUs and parallel programming working under Dr. Miaoqing Huang. He will be responsible for implementing the hardware aspect of this project as well as any other requirement that would need to be implemented.

**Jeff Johnson**  - Johnson is a senior Computer Science major in the Computer Science and Computer Engineering Department at the University of Arkansas. He has completed Database Management Systems, Computer Networks, and Software Engineering, and will share responsibility on most of the project’s software and database implementation tasks.

**Vitaly Borodin -** Borodin is a senior Computer Science major in the Computer Science and Computer Engineering Department at the University of Arkansas. He has completed Algorithms, Computer Challenges, Computer Networks, Embedded Systems, and Software Engineering. He currently has a full-time job at Hewlett Packard developing android applications and working on SQL database. He will be responsible for the SQL server and any other software that needs to be implemented.

**Sarah Hernandez, PhD**. - Dr. Hernandez is an assistant professor at the University of Arkansas in the Civil Engineering Department. She specializes in transportation planning, intelligent transportation systems, and traffic engineering. She received her PhD in Civil and Environmental Engineering with a specialization in transportation systems engineering from the University of California, Irvine. She holds a M.S. from the University of California, Irvine (Irvine, California) and a B.S. from the University of Florida (Gainesville, Florida).

**6.0 Facilities and Equipment** - Jackson Haley

A laboratory will be used in order store the equipment as well as being able to carry out the research and development for the project.

A Raspberry Pi will be utilized to process the mobile device signals. It will receive input from Wi-Fi and Bluetooth dongles connected to it. A LAMP server (Linux, Apache, MySQL, PHP) will be utilized to receive data from the Raspberry Pi, store it, and process/analyze it. The device will be stationed in the University of Arkansas buses, most likely one at a time, and possibly bus stops as well. We will use a Bluetooth receiver to scan for Bluetooth devices and collect their unique and identifying Bluetooth MAC as well as a Wi-Fi add-on to collect nearby device’s wireless MACs.

## 7.0 References

[1] Sarah Hernandez Ph.D., “*Development and testing of a WiFi-Bluetooth System for Transit Ridership Collection*.” University of Arkansas- Fayetteville, 2016.

[2] Pengfei Zhou, Yuanqing Zheng, Mo Li. “*How Long to Wait?: Predicting Bus Arrival Time*

*with Mobile Phone based Participatory Sensing,*” Nanyang Technological University, Singapore, 2011

[3] Bruce Schaller, “*TCRP Synthesis 63: On-board and intercept Transit Survey Techniques*”, Washington, D.C 2005.

[4] G Scott Rutherford, Yinhai Wang, Kari Edison Watkins, Yegor Malinovskiy, “*Perceived and Actual wait time Measurement at Transit Stops Using Bluetooth*”, University of Washington.

.