George Mason University
DAEN 690 | Fall 2019

Traffic Management
Modeling for Manassas Virginia
Sponsored by ASSETT Inc.

Team Members
Goutham Chalamala
Marissa Esguerra
Michael Lepe
James Murray

Table of Contents

Table of Figures

File: DAEN690_ASSETT Final Paper_20191214.docx

Abstract

Vehicle traffic management has become a major issue in the US and especially the Northern Virginia Area. This study assesses the traffic factors and develops models for the enhancement and optimization of traffic flow in the Manassas Virginia area. It focuses on a portion of Prince William County, Virginia with the objective of enhancing traffic flow traffic volume through the area. The secondary objective is to reduce the median traffic transit time for vehicles that go through the area. With these two measures the project team has been provided critical foundational data provided by our sponsor, ASSETT, Inc., from a traffic analysis provided at intersections in the area. The data is also supplemented with VDOT traffic data from their public website, and data from the HERE website https://www.here.com. To assess capacity of the road network, simulated data was also input for random trips in the network.

The George Mason University (GMU) analysis team used their knowledge of data science and data engineering, along with tools for traffic analysis such as Simulation of Urban Mobility (SUMO), OpenStreetmaps, NetEdit, and TraCI. The team developed a series of models to which varied the traffic volume patterns (routes), traffic quantities, traffic light timing and sequencing. The results were analyzed and reported on a 16-hour period by traffic in, traffic out, total flow, and median transit times.

The results provided a series of vehicle profiles for by time of day and on an aggregate as well as at key intersections. These analyses will aid ASSETT, Inc. in their understanding of traffic patterns and constraints.

As additional scope, the simulation development team has proposed a follow-on task with the use the FLOW module which integrates into the SUMO suite of tools to provide a reinforcement learning model of the network. It was assessed that this would be beyond the scope of this course project for the fall semester of the Data Analytics and ENgineering (DAEN) 690 project to complete the reinforcement learning model, but to define the scope and environment with appropriate data sources and include this as a recommended additional research.

1   Introduction

1.1   Background and Rationale

With the increasing populations in the northern Virginia area, and the strong economic advantages of the region with government, technology and a well-educated skill base, the transportation needs of the region are being heavily tasked. The road networks and the transportation systems are struggling to accommodate the population. The influx of population requires additional capacity on roadways established in the Civil War (~1865). As the population increases and the workers need to commute to regional economic centers the traffic congestion begins to build and has caused an increase in commute times, so that the northern Virginia area has a higher than the national average commute time of 26.4 minutes with a value of 39.1 minutes. With overcrowding of roads and global warming issues, storms have become more severe and impact traffic causing large variances in travel time due to delays and accidents. [10,11]

Our sponsor is concerned about the continuing impact of the transportation impacts in their corporate headquarters area in Prince William County, Virginia. The impacts on the employee's timeliness and morale become significant factors in the labor market in northern Virginia. [12] The company is a medium size business with a focus on high technology applications to the environment and economic stability.

With issues such as global warming, due to increased vehicle emissions and traffic road rage incidents due to long commute times and traffic frustration, on the rise in the US, this study provides a modeling and simulation framework and an example instance of one small area of northern Virginia to assess these issues. By developing a prototype example and instance in what has been deemed as the second most congested area in the country, northern Virginia, this approach has great applicability to broader regions and can be a tool for research and regional traffic management. [4]

Vehicle traffic management has become a major issue in the US and especially the northern Virginia area which is our problem space. This study assesses the traffic factors and develops models for the enhancement and optimization of traffic flow in the Manassas, Virginia area, which leads to our project objective. It focuses on a portion of Prince William County, Virginia with the objective of enhancing traffic flow traffic volume through the area. The secondary objective is to reduce the median traffic transit time for vehicles that go through the area.

The George Mason University DAEN course of study addresses global and regional issues in their capstone projects for data analytics and engineering program, this topic was sponsored by ASSETT, Inc., a medium size business in the Manassas, VA. area and was deemed to have significant impact to the region as well as to employ data analytic skills critical to provide a significant impact. Consistent with the ASSETT, Inc. company objectives and business model of applying high technology solutions to practical problems facing the company's stakeholders and their employees. Within this region 75% of households commute to work in the surrounding suburbs of Washington, DC. With few mass transit alternatives and those that do exist being overtaxed, mass transit systems in the area are causing employees to leave for work between 5:00am-10:00am causing a stretch in core hours and to commute longer to their place of employment. The study proposes several means of assessing and optimizing traffic flow in the Manassas area of Prince William County, Virginia. With a success model in this area a broader implementation can reduce carbon emission, reduce transit time, and aid the people of the area to have more time to spend rather than commute for the average 39.1 minutes per day each way to work. The traffic management will

also help the VDOT in understand the best use of traffic highway funding to alleviate transportation bottlenecks and therefore enhance the use of tax revenue to assist the population in their quality of life.

The consequences of not undertaking this project would be:

- Additional impacts to ASSETT employees and delay in productivity due to commute delays
- Potential loss of skills due to transportation delays and lifestyle choices
- Overall increases in employee stress levels and productivity
- Unstructured area development leading to further urban sprawl that is uncontrolled and limiting the growth in the area and the business
- Further congestions and frustration of residents and employers leading to road stress, sick days, and additional accidents or vehicle crimes
- Additional carbon emissions causing urban smog and contributing to global warming
- Critical point potential or current employers leaving the area, having a negative impact on job opportunities,
- Impact on state revenue, funding resources, and opportunities for its residents.

Worker studies have documented the negative impacts on worker morale and employee satisfaction. [13,14]

1.2   Research

Traffic and vehicle engineering are a global issue. With the urgency of global warming as an international issue, world scientists are seeking a diverse number of ways to enhance and optimize our transportation needs and to reduce the overall greenhouse emissions. Traffic management is a key in this endeavor and is a tractable issue within our current technology and processes. With the advent of advanced sensors and intelligent vehicles, several new technology enhancements are on the horizon. These technologies include:

- Cleaner and smarter vehicles which are currently very expensive
- Smarter cars that communicate their path to an overall network controller and manage the flow of traffic across a broader region
- More intelligent traffic lights that will not be preprogrammed and adapt dynamically to traffic needs
- More integrated and intelligent sensors in the roadways and across the region to better coordinate and manage the flow of traffic
- Broader communications networks that better assess traffic trends and mitigate backups

All these alternatives require significant investment and a significant duration of time to implement. The challenge we are facing is the cost and ubiquitous implementation, which may take several years to implement. While there is an urgent need now to "do what we can" in a lower technology approach. It is for this reason the SUMO and a group of commercial traffic optimization platforms were developed.

There are several commercial tools with advanced capabilities to include:

1) **The ONE simulator** was developed at Aalto University and is now maintained and extended in cooperation between Aalto University (Comnet) and Technische Universität München (Connected Mobility) https://akeranen.github.io/the-one
The ONE is a simulation environment that is capable of:
   - Generating node movement using different movement models

File: DAEN690_ASSETT Final Paper_20191214.docx

- Routing messages between nodes with various DTN routing algorithms and sender and receiver types
- Visualizing both mobility and message passing in real time in its graphical user interface
- ONE can import mobility data from real-world traces or other mobility generators. It can also produce a variety of reports from node movement to message passing and general statistics

2) **PTV Visum** is the world's leading software for traffic analyses, forecasts, and GIS-based data management. It consistently models all road users and their interactions and has become a recognized standard in the field of transport planning. Transportation experts use PTV Visum to model transport networks and travel demand, to analyze expected traffic flows, to plan public transport services and to develop advanced transport strategies and solutions. http://vision-traffic.ptvgroup.com/en-us/products/ptv-visum/

3) **The Transportation Analysis and Simulation System (TRANSIMS)**, is an integrated set of tools developed to conduct regional transportation system analyses. With the goal of establishing TRANSIMS as an ongoing public resource available to the transportation community, TRANSIMS is made available under the NASA Open Source Agreement Version 1.3 and is supported by this online community. https://sourceforge.net/projects/transims/

4) **Eclipse SUMO – Simulation of Urban Mobility**
https://www.dlr.de/ts/en/desktopdefault.aspx/tabid-9883/16931_read-41000/
SUMO is a free and open traffic simulation suite which has been available since 2001. SUMO allows modelling of intermodal traffic systems including road vehicles, public transport, and pedestrians. Included with SUMO is a wealth of supporting tools which handle tasks such as route finding, visualization, network import and emission calculation. SUMO can be enhanced with custom models and provides various APIs to remotely control the simulation.

Traffic simulations facilitate the evaluation of infrastructure changes as well as policy changes before implementing them on the road. For example, the effectiveness of environmental zones or traffic light control algorithms can be tested and optimized in a simulation before being deployed in the real world. The simulation platform SUMO offers many features:
- Microscopic simulation - vehicles, pedestrians and public transport are modeled explicitly
- Online interaction – control the simulation with TraCI
- Simulation of multimodal traffic, e.g., vehicles, public transport and pedestrians
- Time schedules of traffic lights can be imported or generated automatically by SUMO
- No artificial limitations in network size and number of simulated vehicles
- Supported import formats: OpenStreetMap, VISUM, VISSIM, NavTeq
- SUMO is implemented in C++ and uses only portable libraries

The state of the art in this area is to extend the modeling and simulations with deep learning and reinforcement learning technologies. The technologies employ predictive analytic technologies with Bayesian Belief Networks or other mathematical models, which identify the principle components of traffic and assess key parameters to identify the causes and impacts of traffic congestion and delay. The German government, along with their aerospace industry, developed an opensource computer tool to help model and assess the complexities of this traffic analysis space. The deep learning tools will allow reinforcement learning to allow a better prediction of the resultant traffic data to the actual conditions and will allow adjustment for it to learn and adapt as the traffic conditions change. The SUMO tool allows us

to propose such an extension which would be accomplished through the FLOW module that integrates to the SUMO suite of tools. [3,4]

> **FLOW is a traffic control** benchmarking framework within the SUMO architecture. It provides a suite of traffic control scenarios (benchmarks), tools for designing custom traffic scenarios, and integration with deep reinforcement learning and traffic microsimulation libraries. https://flow-project.github.io/ [3]

Traffic systems can often be modeled by complex (nonlinear and coupled) dynamical systems for which classical analysis tools struggle to provide the understanding sought by transportation agencies, planners, and control engineers, mostly because of difficulty to provide analytical results on these. Deep reinforcement learning (deep-RL) provides an opportunity to study complex traffic control problems involving interactions of humans, automated vehicles, and sensing infrastructure. The resulting control laws and emergent behaviors of the vehicles provide insight and understanding of the potential for automation of traffic through mixed fleets of autonomous and manned vehicles.

Our GMU development team will explore FLOW in conjunction with Dr. Vadim Sokolov, of GMU School of Engineering, and his doctoral graduate student, Ms. Azadeh Yazdi. The GMU team's objective is to explore the application of FLOW and to scope a follow-on project of enough scope as to be implemented in a following semester.

1.3   Project Objective

As coordinated with our sponsor this project has two project objectives or "objective functions."

Objective function #1 is to maximize the overall capacity of the regional traffic network provided by the sponsor. This is a region of Prince William County, which surrounds the sponsors facility and impacts the employees' transit to work.

Objective function #2 is to minimize the median transit time of vehicles traversing or terminating in the specified region.

With these functions our study assesses: road speed and conditions, traffic light timing, road capacity, and a series of distractors or factors reducing traffic flow. While the customer is interested in traffic "during evening rush", 2:00pm – 6:00pm hours, the model and simulation will assess traffic states against the two objective functions across a 16-hour period.

Our approach provides overall traffic statistics across the road network in total. It will identify traffic entering and leaving at each of the 32 intersections provided by the sponsor for entrance to the road network. This will be done on an hourly bases for a 16-hour period. We also add to output the median transit time for vehicles that have completed the transit of the roadway and indicate how may cars are still within the road network as a residual transit time factor to be considered

1.4   Problem Space

Our team addressed the "Problem Space" as an optimization problem from the operations management domain and recognized the complexity of multiple variables of various levels limiting or constraining the problem and constraining the solution. We recognized that to address these variables, a modeling and analysis tool would be required to help manage the complexity. To address these factors, we performed a search of traffic simulation and modeling tools to help manage the tractability of the numerous variables. We recognized that to solve this problem, a control theory approach would be required and a means for coordinating timing of vehicles via road detectors and traffic light timing.

By utilizing a suitable control policy, a suitably designed urban traffic signal control can decrease problems like overcrowding, stop delay, air and noise pollution, fuel consumption, discomfort and stress. Control policy of traffic control systems are grouped into two main categories, fixed time strategies and vehicle motivate strategies. Fixed time strategies are implemented off-line using maximum efficiency codes based on important and historical traffic data. These strategies don't use information of real time traffic condition. Vehicle motivated strategies carry out an on-line synchronization and optimization of the signal scheduling programs. These strategies are traffic reactive signal control policy utilizing signal timing scheme that automatically answer to traffic requirement.

From a methodological viewpoint, because of different conflicting purposes, serious problems arise when controlling signalized intersections. Firstly, searching for the best fluidity is sometimes incompatible with the best safety. Controlling the traffic from a safety point of view implies constraints on the traffic signal color durations or correlations between traffic signals. These constraints imply limitations on traffic fluidity management. On the other hand, the users of traffic networks such as pedestrian, emergency vehicle, and bicycle often have concern crossing of an intersection. These conflicting objectives require management choices, which will favor certain elements over others.

Simulation of traffic systems using computer software is the most common method adopted these days. More recent methods use either discrete event simulation or continuous-time simulation. Discrete event simulation (DES) is a method of simulating the behavior and performance of a real-life process, facility or system. DES is being used increasingly in health-care services and the increasing speed and memory of computers has allowed the technique to be applied to problems of increasing size and complexity. Traffic lights with single server queues can be modelled using discrete event simulation, as servers are usually at a single location and so are discrete. Many software packages like One, SUMO, Visum, and more generic packages such as Flexsim, Simevents, Promodel, etc., are available for discrete event simulation.

With simulation becoming widespread in traffic engineering practice, questions about the accuracy and reliability of its results need to be addressed convincingly. A major criticism related to this issue is proper calibration of the simulation parameters as well as validation, which is often not done or dealt with in an ad hoc fashion. A complete, systematic, and general calibration methodology is presented for obtaining the accuracy needed in high-performance situations. A technique for automating a significant part of the calibration process through an optimization process is also presented in our approach. The methodology is general and is implemented on a selected SUMO simulator to demonstrate its applicability.

Traffic simulation is a widely used method applied in the research on traffic modelling, planning and development of traffic networks and systems. From the literature study, a variety traffic simulation models were found in experiments and applications with aims to imaginary real traffic operations. The traffic simulation models can be categorized into three namely, microscopic modelling, macroscopic modelling and mesoscopic modelling. This report is aimed to implement a combination of these traffic simulation models, in term of its function, limitation and application within the SUMO framework.

The specific analytic questions to be answered in this report are the objective functions as influenced by the constraints.

1. **Objective Function #1** – To maximize the flow of traffic through the defined network as provided by the customer
2. **Objective Function #2** – To minimize the median transit time for the population in this network

Our objective is to bind an optimal balance between these two objectives and to assess any deviations in the order provided.

The data science scientific methods applied to this issue are:

- Macro-modeling and analysis
- Micro-modeling and analysis
- Traffic engineering analysis of flows and constraints
- Incremental stepwise/piecewise analysis of traffic aggregation
- Stratification of traffic types and introduction into an overall model
- Introduction of stochastic loading parameters to assess capacity limitations and bottlenecks
- Introduction of real-world constraints in traffic modeling

The simplifying assumptions that are being implemented are:

- Application of simplifying assumptions in order to define a more tractable problem (but one that may divorce it from the real world). Some examples: Ignoring right turns, collisions, ignoring certain types of vehicles, pedestrians, etc. Initially no pedestrians or other interference with vehicle traffic.
- No breakdowns or delays
- No emergency vehicles
- Data able to be created randomly on routes defined in the 8 categories
- We have such a small data point that we will simulate data and assess overall volume at key points/ intersections to calibrate the data in a rough manner
- More data is available from HERE, Inc. but it is costly, and we would need to download, clean and assess 4-5 megabytes of data at a significant cost
- We will start with a fixed demand model for vehicular travel and then add variability to the model as time and resources permit to assess variance. See https://en.wikipedia.org/wiki/Induced_demand

With these analyses defined we address a primary user stories to which we apply these models for a real-world concept of operations.

1.5   Primary User Story (-ies):

The primary user stories are centered around the vehicle travel population that transits this area. An analysis of the populations determined that there are 8 major categories or strata of data that comprise the major transit populations that are shown in *Figure 2: Transit Population for the Defined Region Under Study* The average commuter in this area spends 39.1 minutes in commuting to work, the statistics provided by VDOT are presented in *Error: Reference source not found*.

As shown in *Error: Reference source not found*, 75.3 % of commuters drive in cars alone and their average commute is 48% longer than the national average.

The GMU team assessed the eight primary use cases for vehicle traffic in the area and defined them as:
1. Morning traffic commute – from neighborhoods to main roads and to main arteries to Washington suburbs starting at 6:00am with groups leaving hourly 6:00am – 9:00am
   2. Return home commute from work in evening starting at 4:00pm and lasting until 8:00pm
   3. Local Traffic delivery of goods – to be a simulated by a random walk to the neighborhoods and major highways

Commuting in Prince William County, Virginia

The typical American commute has been getting longer each year since 2010. The average **one-way commute in Prince William County takes 39.1 minutes**. That's longer than the **US average of 26.4 minutes.**

How people in Prince William County get to work:
- 75.3% drive their own car alone
- 12.3% carpool with others
- 4.2% work from home
- 5.3% take mass transit
- 3.2% other

Figure 1: VDOT Statistics on Prince William County

4. School bus traffic – similar to local traffic but moving more slowly and gating or slowing traffic with a max speed of 30 mph and stops evert mile. This will be a special type of vehicle.
5. Daily shopping by homemakers or work at home residents – will go to stores and or malls. In local area from 10:00am – 2:00pm – normal car traffic. Could be implemented by randomized routes in the local area
6. Crossing traffic through the area for commute and reverse commute times – will coincide with commutes but will only use major routes through the area time coincide with commute and reverse commuter times. See items #1 and # 2 above for timing.
7. Prince William Regional Transit Company (PTRC)/ Washington Metro Transit Area (WMTA) Transit Bus traffic for industries in the area and shops – primarily go to malls or major industry the bus will travel at 30 MPH and stop evert ¼ mile
8. Commuters that come to the area for business and work. These are primarily located along major routes and are identified as manufacturing on the map

1. Morning traffic commute leave suburbs to go to highways at 6:00, 7:00, 8:00, 9:00 and   This process is reversed in the evening when the commuters come home.
2. Return home commute in evenings at 4:00, 5:00, 6:00, 7:00, 8:00 + 39-minute one-way commute.   This route is the reverse of the route #1 with a different time and in reverse order
3. Traffic for delivery drivers during the day / UPS/ FedEx make random deliveries   Similar to bus traffic for loading of local roads and deliveries

4. School bus traffic for:
    Elementary students go in at 9:00 and come home at 4:00pm
    Middle school at 7:45am and home at 2:20pm
    High school at 8:20am and home at 3:00pm

5. Daily shopping and activities – to stores and malls or businesses – people shopping during the day and running errands to businesses.

6. Crossing traffic to get to work for the 39-minute commute – crossing the area – presumably on the major roads
    To transit the area on a major road takes 20 minutes so there is another 20 minutes of commuting one way – see where people go. – Chantilly, Reston, Springfield, Tysons, Leesburg
    Take Rt 66, Rt 28, Rt 234 – via major purple routes as noted on map (this would happen in AM and reverse in PM during "rush" hours.)

7. Business traffic and major businesses that draw traffic for local companies and people working in the area like the GMU Prince William Campus, companies in the area
    Others – investigate by size of parking lot or parking garage – develop stochastic model (may be too detailed)
    Use google maps to get rough size of parking at facilities and estimate car-load.

8. Other routes feed the main routes due to area segmentation with natural boundaries or barriers.

Figure 2: Transit Population for the Defined Region Under Study

The percentages of these populations will need to be varied to develop a composite traffic load for the area. These variations performed to adjust for different times of day. They represent the single independent variable that most affects the relative percentages of these eight groups, i.e. the percentage of drivers who are commuting to/from work is very high at 7am, but low at noon or at midnight. Likewise, school buses make up a large portion of traffic at 3pm, but not much at 7pm. There are no firm statistics on the percentage of vehicle drivers that take these routes, but they were varied in accordance to the data in Figure 3: Traffic Volume Estimates for Simulations below.

Major Scenarios / Strata of traffic considered		Min %	Max %
1. Morning traffic commute		60%	85%
2. Return home commute from work in evening		Opposite of above	Opposite of above
3. Local Traffic delivery of goods		6%	10%
4. School bus traffic like local traffic		8%	8%

5. Daily shopping by homemakers or work at home resident's area		8%	8%
6. Crossing traffic through the area times		5%	5%
7. Washington Metro Transit Area (WMTA) Transit Bus traffic		3%	3%
8. Commuters that come to the area for business and work		10%	10%
Number may exceed maximum		100%	129%
Stochastic distributions - Gaussians are added for simulated volume			

Figure 3: Traffic Volume Estimates for Simulations

This project has the potential to inform the population as to smarter routes to take like the "waze" web application https://www.waze.com and to help enhance and inform the driving public. This will enable additional intelligence and projections to be incorporated in a predictive nature to such an application. This will also help business and delivery personnel to have better "on time" prediction to enhance productivity reduce wasted time in traffic and increase efficiency.

This set of user stories explicitly state what the project is attempting to address. Our GMU DAEN-690 project is focused on our ASSETT, Inc. customers' requirements and is based on the user context and value proposition, we developed the following primary user story to guide our project:

> *"As a User, I want to reduce the transit time through the area and provide a predictable and reliable experience to enhance the quality of life of the residents and businesses prosperity in the region."*

1.6   Solution Space

There were three major possible approaches to addressing problems in this solution space. They are:

1. **Construct a specific discrete simulation** in a tool such as Flexsim, Simevents, Promodel, Matlab, Modsim, or Simulink. This would be a very custom and intensive method or coding, data assessment and analysis and tuning. Several algorithms and procedures would need to be recreated.

2. **Use a traffic simulation tool** as discussed earlier in Section 1.2 Research of this document. The tools assessed were: ONE simulator, PVT Visum, TRANSIMS, Eclipse SUMO, and a few other approaches. This type of tool incorporates the traffic management concepts and provides menu driven parameter specification. This type of tool includes models and iconography that relate to the traffic simulation domain and are controlled though easy to implement XML interface files.

3. **Deep Learning or Reinforcement Learning models-** this method requires significant data in a supervised (curated or marked valid and invalid) or unsupervised manner raw data that is supplied to the simulation and the neural network will establish weights or parameters for determining a projected outcome. This approach is a "state of the art approach and would require reinforcement learning to adjust or "tune" the deep learning model. This technique is generalizable to a large area and is in the initial assessment phase by researchers. It was determined by the GMU Project Team and Dr. Vadim Sokolov, our technical advisor, that this was beyond the scope of our one semester project. So, it was suggested that we propose a project plan for a follow-on project to consider that approach once approach #2 would provide a foundational analysis.

In the resulting trade-off analysis, it was requested by the customer ASSETT, Inc. that we develop a model in a traffic simulation tool with a vehicle simulation technology model. The data required for a

deep learning or reinforcement learning model would need to be obtained from the HERE data base and would require additional time and cost that the team was not able to accommodate. The sponsor, ASSETT, Inc. was very generous in providing data and performing an in-person survey of the area. However, the data was limited to times on one day in the evening rush and was to sparse and discrete with a timeframe which our overly constrained a model. The data was used for calibration and bounding by needed to be augmented with other synthetic data from VDOT projections and other environmental assessments. *Figure 4: Compare and Contract Possible Approaches to Traffic Simulation* represents the tradeoff comparison and contrasting the pros and cons of each method, both from the perspective of feasibility/risk/analytical accuracy and from the perspective of the eventual user experience. The resultant matrix depicts the possible approach that our team took to solve this in approach selection.

Approach Evaluated		Criteria		Pros	Cons
Construct a specific discrete simulation					
	Criteria	H	Can be very specific and specialized as a custom-built solution in a tool	X	
	feasibility	L	Requires a lot of time with the tool and coding		X
	risk	H	very time consuming and computer intensive. Requires custom visualizations	X	
	analytical accuracy	H	most accurate	X	
	User experience	L	Not as good as Traffic Simulation		X
Use a traffic simulation tool					
	Criteria	H	Able to be accommodated by parameter files	X	
	feasibility	H	Possible within time duration	X	
	risk	M	Need to manage computer runs and # of permutations	=	=
	analytical accuracy	H	Able to be tuned with sensors/detectors but requires a detailed plan	X	
	User experience	H	Visualizations are highly applicable to the traffic domain	X	
Deep Learning or Reinforcement Learning models					
	Criteria	L	curated		X
	feasibility	L	Extensive modeling requiring tight timelines		X
	risk	H	Unlikely to be done in a semester	X	
	analytical accuracy	H	High and generalizable to other areas	X	
	User experience	M	Not as good as Traffic Simulation	=	=

Figure 4: Compare and Contract Possible Approaches to Traffic Simulation

In conclusion the use of the Traffic simulation tool (SUMO, NetEdit, TraCI) met many of the criteria and was consistent with the sponsors expectations. To accommodate the more advanced data mining and reinforcement learning techniques, our team decided to provide a section of the report on the approach to deep learning to address the more advanced technological and academic implementations for future consideration and a possible future project.

Our system delivers value to its users when it accurately reports projections on traffic flow and transit time so that the user can better navigate the area and arrive within the expected time with potentially reduced stress. Users derive value from these estimates when they feel more confident in their chosen approaches or avoid being misled when left to their own uninformed assessments or are presented with inaccurate data. We expect our system can help guide users to more authoritative repeatable and predictive routes for traffic management and assist business outlets in more efficient delivery by altering routes and delivery methods.

1.7   Product Vision

Our customer's vision is to provide value that exceeds the ever-growing needs of today's cities. They want to build the next generation of mobility solutions on the world's most complete location platform. We recognize that not all stakeholders will directly use the results of this analysis, but the stakeholders may be further informed of the factors, this will be a follow-on action by our sponsor. The group which will interface directly with this analysis will be the drivers in this region. If the sponsor so chooses this could be used as an amicus brief to support a position for road improvements or to lobby for road improvements in the region. This independent analysis can be used to confirm or refute other analyses from VDOT, or government and/or planners.

By doing so they will empower your city's mobility services with location intelligence to:

1) Reduce traffic congestion
2) Improve transport planning
3) Create consistent digital UX (user experience)
4) Ensure frictionless end-to-end journeys

A description of the new user experiences with this product are realized in the following ways:

•**For:** Vehicle drivers or passenger of cars, busses, trucks, or almost any transportation vehicle

•**Who:** Travel by car or bus in the US or in the developed or developing world

•**The:** model or simulation can reduce stress and provide a predictive assessment to help manage expectations for on time delivery and provide business intelligence for companies to increase competitiveness

•**Is a:** model or simulation that is a projection for planning purposes

•**That:** allows expectation management and allocation of resources for business optimization

•**Unlike:** more technological solutions that may take several years to evolve and cost millions of dollars this solution is here today and can evolve with new technology providing a foundation for the future

•**Our product:** through practical modeling and simulation approach will reduce travel time by reducing congestion and enhancing the user experience

These product vision statements are realized through an applied scenario analysis in the following four (4) short scenario summaries. These scenario summaries help to explain the value provided through this product to the various customer segments so that the vision can be better understood by their valued customers.

Scenario #1 - Reduce traffic congestion

- **The road to efficiency**. Pair the right vehicle with the right delivery at the right time and optimize order stacking. Move your fleet seamlessly through urban environments with dynamic routing and real-time location technology
- **Strengthen user loyalty.** Ensure on-time delivery, reduce wasted trips and meet local delivery demands with rich, fresh and accurate location intelligence. With precise drop-off locations and real-time ETAs, customers become repeat users.

- **Data-driven profitability**. Capture and analyze data to find optimization opportunities. Overcome weather disruptions, traffic delays and shifts in customer demands with comprehensive data analysis
- **Prepare for new services -** Ready for the global food fight? Find out how location services can help online food delivery companies stay competitive in a global industry worth over $94 billion – and growing
- **Caveats:** This implementation requires the collection of traffic parameters and performing a meaningful assessment or analysis of the parameters, by a realistic set of parameters which need to be used for realistic assessment such as: pedestrian traffic, emergency vehicle traffic, traffic accidents, road closures, road work, etc.

Scenario #2 - Improve transport planning

Drivers lose an immense amount of time and money navigating through traffic. In tandem, governments and businesses struggle to manage growing congestion levels and related safety risks. Intelligently predicting and reacting to changing road conditions is essential in solving these problems.

This approach will aid drivers with access to more accurate information about congestion and incidents. Information such as construction zones, accidents, traffic routes and estimated time of arrivals. Real-Time Traffic delivers this precise information to create a smarter and safer driving experience, wherever the destination.

Scenario #3 - Create consistent digital UX (user experience)

Unmatched value for better road travel. This approach offers innovative traffic features such as Traffic Safety Warning, which alerts drivers of an incident with detailed live traffic information and re-routes them so they can reach their destination safely.

Available from your:
- Smart phone,
- Vehicle guidance system
- Office locations for information exchange and location
- Customer locations for time and arrival
- For many uses such as:
    - Service delivery time estimates and route planning
    - Situational awareness for emergency, breakdown or timing
    - Coordinated service delivery

Scenario #4 - Ensure frictionless end-to-end journeys

Provide a seamless user experience. Develop personalized and customizable applications based on the needs of your users using location intelligence. Strengthen customer loyalty by providing new services, a consistent digital experience, and timely offers.

Optimize business operations. Use applications, platforms and mapping content to scale and improve time to market. Leverage sophisticated technology for fleet planning and trip analysis to help with cost reduction and service efficiency.

Maintain a competitive edge. Future-proof your mobility services with new capabilities, including big data processing and data visualization. Ensure your operations are secure and flexible by leveraging one-stop development platform for delivery applications which will:

- Optimize time

- Reduce stress
- Reduce cost of travel
- Enhance the experience
- Enhance predictable schedules or provide options
- For driver, passengers, cargo
- Increase predictability for delivery

1.8   Definition of Terms:

In an effort to assist the reader in better understanding technical and academic aspects of this report out team has included a short definition of terms section to aid the reader. Below is a short list of terms and their definitions for reference as needed

- ASSETT – Company name as the sponsor of this project
- DAEN – Data Analytics and Engineering – the department in GMU under which this program was administered
- DES – discrete event simulation – a type of computer model developed from discrete events
- DTN routing – Deterministic Routing Network – a set of routing protocols
- ETA – Expected Time of Arrival – the anticipated time that an event will occur
- FLOW – a reinforcement learning module developed by UC Berkeley for simulations
- GIS – Geographic Information System – a map-based system or product
- GMU – George Mason University – the name of the university where this study was produced
- HERE - https://www.here.com/products/location-based-services/
- OSM – Open Street Map a commercial sources of map data www.openstreetmap.com
- Jam Factor – a measure that calculates the "free flowing"-ness of a road. Used in the HERE dataset, with a value of 0.0 being "free-flowing" and a value of 10.0 being "road closed".
- RL – Reinforcement Learning – a type of deep learning or machine learning that provides feedback to the system to allow it to modify its algorithms and to "learn" as it progresses
- SUMO - Simulation of Urban Mobility – an open source simulation tools developed by the DLR in Germany
- TraCI - TraCI is the short term for "Traffic Control Interface". Giving access to a running road traffic simulation, it allows to retrieve values of simulated objects and to manipulate their behavior "on-line". https://sumo.dlr.de/docs/TraCI.html#introduction_to_traci
- TRANSIMS - Transportation Simulator – used in transportation systems
- VDOT – Virginia Department of Transportation
- VISUM, VISSIM, NavTeq – map formats
- WMTA – Washington Metropolitan Transportation Authority
- XML – eXtensible markup language

2 Data Acquisition

2.1   Overview:

The data acquisition process of this project was focused upon the HERE data from
https://developer.here.com/documentation/traffic-data-service/topics/concept-realtime-traffic-data.html
the site provides Real-Time Traffic Data as stated on their website:

> "HERE Real-Time Traffic comprises highly accurate map data from multiple sources, including connected car probes, roadway sensors, and live operations centers.

> *Real-Time Traffic can be used to ascertain current traffic conditions for map display and route guidance, including alternative routing when a traffic incident is detected on the planned route.*
>
> *Real-Time Traffic contains two types of traffic data:*
>
> - *Flow - this is information about the speed of travel and travel times along a segment of a roadway.*
>
> - *Incident - this is information about events that are affecting the flow of traffic or that may be important for drivers to know.*
>
> *The HERE Traffic Data Service supports Real-Time traffic data in the US and Europe. [N]"*

The services support traffic data in the United States and in Europe the service offers three level of service

- Freemium - $0/month but limited access to
    - 250K Transactions per month
    - 5K SDK Monthly Active Users
    - 250 Assets per month
    - Pay per additional Transactions
- Pro - $449/month
    - 250K Transactions per month
    - 5K SDK Monthly Active Users
    - 250 Assets per month
    - Pay per additional Transactions
- Premier
    - Increased usage limits

Given that there are no additional funds the Freemium service was used. ASSETT, Inc. did offer for the university to download data with their script, but no dollar amount was assessed; therefore, we felt it would not be in the best interest of the project to experiment and to drive up a potentially large bill.

In addition to the HERE traffic data, ASSETT, Inc. provided an additional dataset which contained intersection information. This dataset was manually collected by the ASSETT team and included different intersection attributes such as light timing, road segments, and direction.

Overall, the project team used a combination of the HERE traffic data and the ASSETT manually collected data, in addition to the preliminarily researched bounding criteria from VDOT, to understand and gauge the volume of the network.

2.2   Field Descriptions:

2.2.1 ASSETT Collected data

The sponsor ASSETT Inc. also conducted a traffic survey and sent representatives of three different teams to the various traffic intersections and gathered data on traffic volumes at the intersections. This data was made available for each of the 32 traffic segments in the desired road network to be studied. The ASSETT, Inc. team gathers several data sets and they consist of:

- Traffic segment – Begin #, End #, Length, Travel Time (s), Speed and Signs
- Traffic at road intersections-
    - Intersection name (junction of the two roads intersecting)

- o Traffic Light Present – Y/N
- o Northbound
  - Left turn # of lanes
  - Right turn # of lanes
  - Regular though lanes
- o Southbound
  - Left turn # of lanes
  - Right turn # of lanes
  - Regular though lanes
- o Eastbound
  - Left turn # of lanes
  - Right turn # of lanes
  - Regular though lanes
- o Westbound
  - Left turn # of lanes
  - Right turn # of lanes
  - Regular though lanes
- o Northbound – Subcycle # - 1 or 2
  - Time
  - Straight
  - Turn L
  - Turn R
- o Southbound – Subcycle # - 1 or 2
  - Time
  - Straight
  - Turn L
  - Turn R
- o Eastbound – Subcycle # - 1 or 2
  - Time
  - Straight
  - Turn L
  - Turn R
- o Westbound – Subcycle # - 1 or 2
  - Time
  - Straight
  - Turn L
  - Turn R
- o Overall time – for traffic signal to cycle
- o Misc. Notes and Comments

2.2.2 HERE collected data

The ASSETT Inc. team pulled data from HERE, Inc. Developer Website using latitude/longitude bounding box to filter the data with a script run once per minute for a week's duration. The result as a set of 16,296 .json files which were converted to .CSV, then the data was combined from each .json file into ONE .CSV. *Figure 5* below depicts the process used:
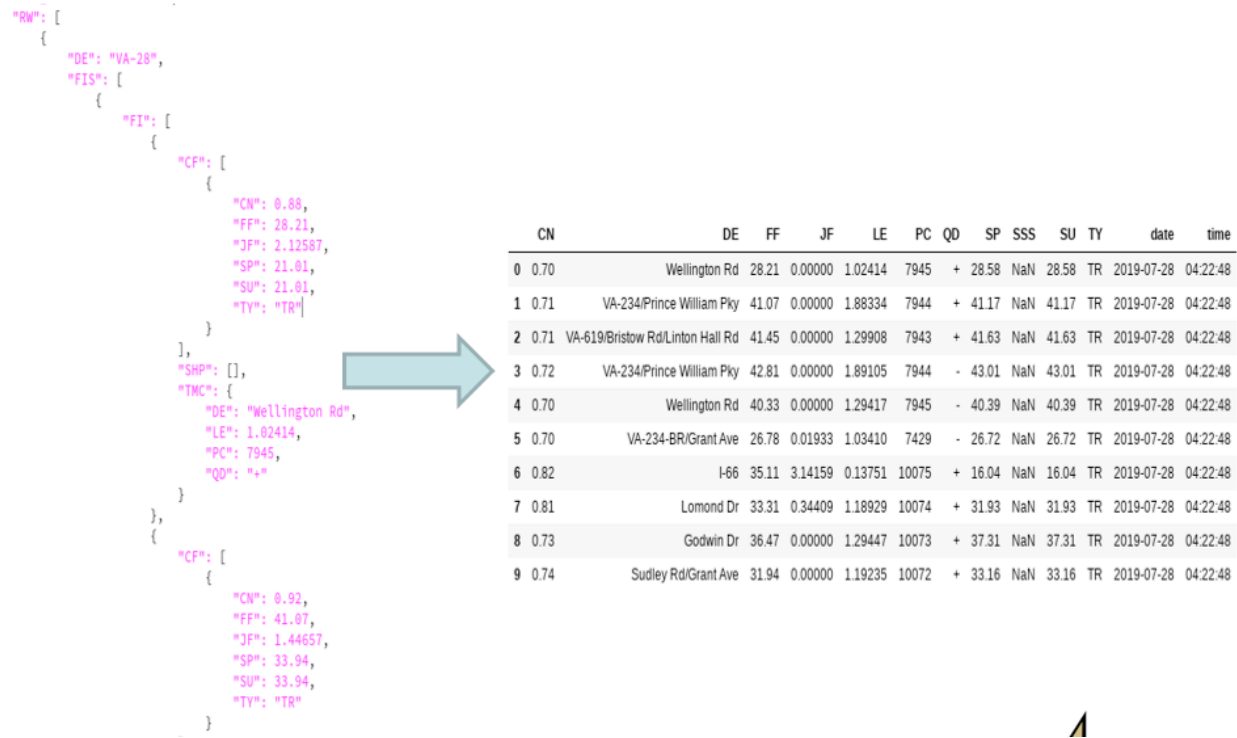
"RW": [
    {
        "DE": "VA-28",
        "FIS": [
            {
                "FI": [
                    {
                        "CF": [
                            {
                                "CN": 0.88,
                                "FF": 28.21,
                                "JF": 2.12587,
                                "SP": 21.01,
                                "SU": 21.01,
                                "TY": "TR"
                            }
                        ],
                        "SHP": [],
                        "TMC": {
                            "DE": "Wellington Rd",
                            "LE": 1.02414,
                            "PC": 7945,
                            "QD": "+"
                        }
                    },
                    {
                        "CF": [
                            {
                                "CN": 0.92,
                                "FF": 41.07,
                                "JF": 1.44657,
                                "SP": 33.94,
                                "SU": 33.94,
                                "TY": "TR"
                            }
                        ]

	CN	DE	FF	JF	LE	PC	QD	SP	SSS	SU	TY	date	time
0	0.70	Wellington Rd	28.21	0.00000	1.02414	7945	+	28.58	NaN	28.58	TR	2019-07-28	04:22:48
1	0.71	VA-234/Prince William Pky	41.07	0.00000	1.88334	7944	+	41.17	NaN	41.17	TR	2019-07-28	04:22:48
2	0.71	VA-619/Bristow Rd/Linton Hall Rd	41.45	0.00000	1.29908	7943	+	41.63	NaN	41.63	TR	2019-07-28	04:22:48
3	0.72	VA-234/Prince William Pky	42.81	0.00000	1.89105	7944	-	43.01	NaN	43.01	TR	2019-07-28	04:22:48
4	0.70	Wellington Rd	40.33	0.00000	1.29417	7945	-	40.39	NaN	40.39	TR	2019-07-28	04:22:48
5	0.70	VA-234-BR/Grant Ave	26.78	0.01933	1.03410	7429	-	26.72	NaN	26.72	TR	2019-07-28	04:22:48
6	0.82	I-66	35.11	3.14159	0.13751	10075	+	16.04	NaN	16.04	TR	2019-07-28	04:22:48
7	0.81	Lomond Dr	33.31	0.34409	1.18929	10074	+	31.93	NaN	31.93	TR	2019-07-28	04:22:48
8	0.73	Godwin Dr	36.47	0.00000	1.29447	10073	+	37.31	NaN	37.31	TR	2019-07-28	04:22:48
9	0.74	Sudley Rd/Grant Ave	31.94	0.00000	1.19235	10072	+	33.16	NaN	33.16	TR	2019-07-28	04:22:48

Figure 5: Representative Process Used for Data Collection And Aggregation

The HERE data consists of 765,818 rows and 14 columns which contain data on 18 main roads and are organized by Point Code (PC) – which is a road segment between two intersections. Relevant columns and their location in *Figure 5* are shown above:

- Road name – column DE
- Jam Factor – column JF
- Queue Direction – column QD
- Date – column "Date"
- Average Speed (mph) – column SP
- Free Flow Speed (mph) – column FF
- Length of street – column LE

The HERE data has a confidence level of all data collected of 70% in real time. However, the data:

- Does not contain data on traffic signals
- Does not contain data in every intersection light cycle
- Average speed and jam factor

It was anticipated that the ASSETT, Inc. provided HERE data exploration report [15] would provide insight for:

- Comparison of average speed
- Investigation of traffic flow in different queue directions
- Comparison of jam factor at different times
- Find differences between predicted speed and actual speed

These factors while promising proved to be limited and only marginal use in the analysis approach determined. These factors may be useful with a significant purchase of HERE data to feed an Artificial Intelligence Analysis (AI) which is assessed as a follow-on task to this project; see Future Work DAEN 690 - Follow on Research in Traffic analysis and Simulation.

2.3   Data Context:

The approach chosen for work with SUMO which requires a knowledge of traffic Routes. These routes are not evident from the data only traffic flow and volumes at key location points such as intersections.

The team analyzed the traffic patterns of the residents in the area a developed the eight (8) general routes traffic follows in the area as noted in Figure 3 above. We would then provide stochastic loading of the road network to meet the limits specified in the HERE data at key location intersections. While this approximates the data a stochastic Gaussian distribution of data was assumed. Other distributions were also assessed and would be tried as time permits. A bimodal distribution, characteristic of a commute day of the week with peaks at 9:00am and 5:00pm. Or another distribution of the triangular shape which is more characteristic of a weekend with traffic building to 12:00 noon and then decreasing until 9:00pm.

The HERE data and its characteristics were used as a bounding analysis and a calibration of the model to develop the appropriate volumes of data at the key points. We realized that it would take a significant amount of data to assess the overall characteristics of the region for a 24/7 assessment and characterization. The VDOT data is also assessed with a 70% accuracy [15] for real time data. Therefore, our model is a reasonable approximation of the flow of data along the various routes. It was determined that the route planning and its general percentages may be a significant insight and byproduct of this study. This is data which is not available at this time and may be available in the future as advanced applications beyond WAZE come on-line but also present Proprietary Information Issues (PII) with drivers sharing personal proprietary data.

After a thorough analysis of the data for this project it was determined that the needed data could not be purchased from HERE Inc due to project budget limitations of no additional funds. Therefore, we decided to use the small amount of existing data from HERE Inc. and the manual data collection as a calibration and bounding analysis and to ensure an appropriate profile of data and to use stochastic data in the formats for gaussian, triangular, and purely random distributions.

2.4   Data Quality Assessment:

The data conditioning performed was conducted by the development of "download scripts" from the HERE, Inc. system by Mr. Nicholas Bruno of ASSETT, Inc. Mr. Bruno provided the initial data and the scripts through our partnership with our sponsor and was instrumental in data explanation and survey performance. The entire team appreciates the effort and support or Mr. Bruno and the ASSETT Inc. sponsor.

The Sponsor, ASSETT Inc reported the following facts to our team relating to report HERE Data Quality.

- HERE shows a confidence factor for all data collected
- All rows above 0.70 confidence (indicates real time) – VDOT reports real-time data is .85 confidence due to system constraints and operational issues
- Does not contain data on traffic signals
- Does not contain data on every intersection and light cycle
- Contains more data than our ASSETT collection

- Average Speed and Jam Factor could prove useful (used only to assess flow of traffic through the region)
- How can we integrate this with collected data?
  - It was used to correlate and to assess correlation of data from the HERE data and the manually collected data. This is used as a bounding analysis to assess and develop traffic profiles and factors as pointed out by VDOT as key factors in traffic analysis and capacity management.

2.5   Data Sources Assessment

The data quality from the HERE Inc. system are stated to be 70% accurate in real time. The data from VDOT are also stated as 70% accurate in aggregate over a 1-year period. Smaller periods are not identified as to accuracy or data quality.

Stochastic data/ synthetic data was used for loading of traffic to the bounds identified in the HERE and ASSETT Inc. data and to assess the overall loading of the road network to project volumetric trends and bottlenecks. In several cases the thresholds were removed to assess overall traffic volumes and to assess saturation points or points of significant variance between median weight times for the drivers. This later indicator is a sign of road saturation as well.

The factors to be assessed in this section include:

- Completeness: The data provided is a subset of one day's data and is limited in time from 4:00pm – 6:00pm on 07/28/2019
- Uniqueness: the data has unique characteristics for traffic management purposes. Not all the data types and fields are to be used. The SUMO product uses OpenStreetmaps as a reference and map of routes. Each road, or portion of a road segment, is labeled with a unique identifier in OpenStreetmaps and the NetEdit program is used to identify road segments for the initiating and terminating of routes and identified in the TAZ files (XML files).
- Accuracy: as stated earlier the HERE and VDOT data is ~70% accurate. The ASSETT Inc. data is highly accurate but only for a short period of time from 2:00- 6:00pm on a single day. Therefore, it is of limited scope.
- Atomicity: Yes, the data is atomic
- Conformity:
  - Due to text encoding, HERE advises to view data in Notepad ++ or open it in Excel by loading text data and setting encoding as Unicode UTF-8.
  - **RDF** - RDF is a standard that provides a common reference model for organizations involved in the creation, update, supply, and application of referenced and structured road network data. RDF enables road network data providers to use a common data publication reference model for their product definition, while allowing users to define their application independent of a specific data provider's model. The RDF is designed to support several application areas, such as Vehicle Navigation Systems, Highway Maintenance Systems, Road Transport Informatics, and Advanced Road Transport Telematics.
  - **Shapefiles** - data is also available - not applicable for this simulation
  - **3D landmarks** – data is also available – not available for this simulation
- Overall quality: The overall data quality is acceptable but limited in scope with the constraints of not purchasing other HERE data and recording routes of traffic which are all notional.

3   Analytics and Algorithms

"Simulation of Urban MObility" (Eclipse SUMO) is an open source, highly portable, microscopic and continuous road traffic simulation package designed to handle large road networks. SUMO is licensed under the Eclipse Public License V2. "Eclipse SUMO" is a trademark of the Eclipse Foundation. If you use SUMO, please support its development by telling us about your Publications. The content of this Documentation is freely editable according to the wiki style.

The SUMO product uses OpenStreetmaps as a reference and map of routes. Each road, or portion of a road segment, is labeled with a unique identifier in OpenStreetmaps and the NetEdit program is used to identify road segments for the initiating and terminating of routes and identified in the TAZ files (XML files). The OD.XML files are used to specify load or volume of traffic for the network route.

The process that we followed is represented in *Figure 6: Map Input and Capture of Real-World Road Networks* and *Figure 7: Process for Conversion and Integration of Input Files to the Simulation* below. This data feeds the configurations and provides a foundation for the required analyses.
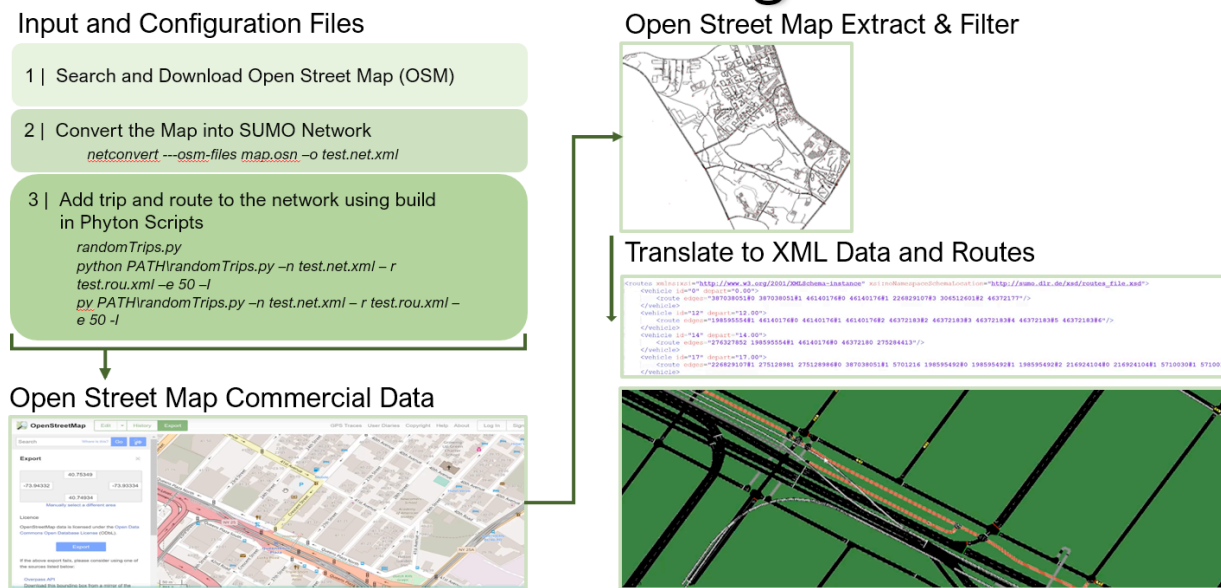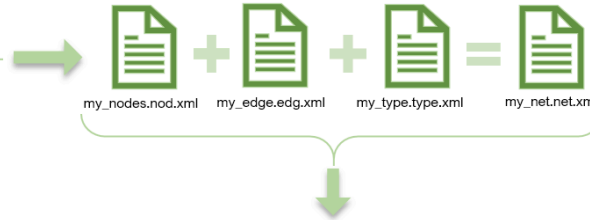

Figure 6: Map Input and Capture of Real-World Road Networks

Input and Configuration Files

Figure 7: Process for Conversion and Integration of Input Files to the Simulation

The data provided by the sponsor was a small sample that served as a point estimate for a 7/28/2019 and for 3 hours within that day. This amount of data is not enough to characterize the traffic profile. The Team could have ordered additional data from the HERE, Inc, database but the data feed and other charges were not able to be accommodated within this project cycle for the DAEN program. Our solution was to use the small point data and to determine an analytical and volumetric profile of traffic from the VDOT and Northern Virginia Data based on transportation and population statistics [10,11,12].

The volumetric profile observed was then extended with characteristic data from VDOT and Prince William County population statistics. A stochastic profile was explored according to national averages and regional specific characteristics that could be correlated with the point estimate data.

Our teams challenge is to estimate the percentage of route traffic available, so that the aggregate data will sum to the calibrated HERE, Inc. and VDOT data for the intersections in the specified area. The balancing of the traffic along with the specification of parameters for the traffic light timing or delay will provide insights to the flow of vehicles within the network. "Detectors" are placed at various locations within the network (at key intersections) to capture traffic counts for specified periods. We also placed "detectors" at the entrance and exit to the road network on the 14 major intersections to capture Inflow and Outflow of vehicles. From this data we were able to capture gross flow of traffic across the network, and the mean time for transiting the network. We were also able to assess specific volumes of traffic at specified intersections.

Our stochastic generation process will then be able to generate and quantify the load or volume of traffic by specific route and route type. In this way we will vary the composite makeup of the traffic to achieve the total expected volume and to vary the conditions based of other projected scenarios as proposed by VDOT or as estimated by the project team. We realize that we are estimating route statistics and grossly estimating composite traffic. The data used in based on traffic analyses from other major cities and data provided from VDOT and Prince William County Census. Our objective is to find and characterize the inflection points in the traffic volume and timing to inform our sponsor. Correlating to a specific traffic load or volume while interesting is not a quality measure that is driving the simulation. Understanding load and road capacity under various loading and commuting scenarios will help to provide insight into ways to adjust policy, procedures, scheduling of traffic lights to better increase traffic flow (throughput).

The SUMO tool routes traffic based on the shortest "Manhattan" route from start location to destination. Each of the scenarios will have start and destination locations specified and routes calculated automatically by the duarouter component application (inside SUMO).

A challenge for the traffic simulation will be to assess the internetwork traffic volume and account for it since, it may not exit the network due the cyclic nature of the traffic strata; for instance, school busses or delivery vehicles. The vehicle tends to take routes within the network gating or reducing traffic flow while rarely exiting the network to be counted. School bus and city bus traffic then to "jam" or reduce traffic due to frequent stops.

Once the traffic routes are determined the traffic-flow will be calculated by assessing the road segment speed limit and the volume of traffic moving through the network element as modified by the traffic light timing or schedule which can be varied. We can vary the traffic light timing accordingly as a key parameter to be assessed. The SUMO and NETEDIT tools allow a menu driven specification of the parameters or a version that allows specification in the XML files with the addition of random distributions.

The data collection and baseline calculations will be performed in Microsoft Excel from each of the computer runs which are required. The number of runs to have a statistically accurate sample is beyond the scope of this study of several runs per scenario. Therefore, this project will only be able to provide initial assessment and a gross understand of the traffic characteristics. We developed a stratified approach to data analysis and segmented the data according to the categories identified in *Figure 2* above. We developed a combination of the various scenario strata data which enabled us to use a systematic approach to data generation. The following matrix, in *Figure 8*, shows the allocation of the type of traffic volume in the appropriate strata allocated to time periods where the data would eb applicable. This enables a systematic buildup of traffic from individual components. Of course, as more components are added together, that is an additional residual delay which is a result of the building backlog of traffic.

	Type of Traffic	Traffic volume Approximate %	5:00	6:00	7:00	8:00	9:00	10:00	11:00	12:00	Time 13:00	14:00	15:00	16:00	17:00	18:00
1	Morning commute	30	X	X	X	X	X									
2	Morning crossing traffic	6		X	X	X										
3	Local Commute	7			X	X	X	X					X	X	X	X
4	School bus traffic	6			X	X	X	X								
5	Delivery Traffic	8				X	X	X	X	X	X	X	X	X	X	
6	Daily shopping	7						X	X	X	X	X	X	X		
7	Evening commute	6										X	X	X	X	X
8	Evening crossing traffic	30											X	X	X	X
		100	1	2	4	5	5	4	2	2	2	3	5	5	4	3

Figure 8: Matrix of Applicable Scenario Related to Type of Traffic

The compounding of traffic is assessed in a few benchmark simulations runs that develop a backlog of traffic volume which takes approximately 1 hour to work-off. Therefore, once the traffic volume reaches 40% an additional delay of 10% is added and with a volume of 60% a delay of 20% is added.

Our GMU team's analysis considers factors in the analytical process that can modify the transit criteria which is the major factor for the customers objective function. They are:

- Road network configuration due to bottlenecks or alternate routes available

- Road size/constraints in key bottleneck areas and alternative

  o   adding traffic lanes at key points

  o   using breakdown lanes in key times

  o   providing express lanes to enhance volume flow

- Identify key locations for mass transit evolution of transitions

- Traffic light timing and flow analyses

- Traffic restrictions or time, location, vehicle type, etc. impacts for specific key periods

- Data collection and profiling for interface to future technology for in vehicle navigation systems

- Data collection for overall vehicle profiling and alternate route analyses

The Analytics effort for this project required the development of over 150 configuration files for maps, detectors, routes, networks, vehicle volumes, signal timing, and options. The debugging effort for these files was significant in the start and converged to a focused analysis to develop the bounding data and analytical products. The final analysis required greater than 50 computer runs for data generation and refinement. The data collection and aggregation required data conditioning to convert data from XML format to CSV format and then to correlate the data by location into appropriate fields relating to map locations. The processing was performed with a Windows 10 Core i5 processor and at least 8GB RAM (32GB preferred) and 200MB of storage for file storage. The segmentation approach to the processing was key in terms of reducing the size and scope of the analyses to allow a common method for compilation of data in a systematic manner. The specification of the files took a long period of approximately 1/3 of our project duration or ~5 weeks. Due to the DAEN 690 program funding profile, no external funds were available for data or external Amazon Web Services (AWS) processing, so the project work was performed with Open Source software and on GMU Team members personal laptops. It is expected that this will evolve with the program maturity and as needs arise. The final analyses required 36 hours to run the base analyses and another 36 hours for data correlation, aggregation, conditioning, and visualization.

Our summary analysis provided an:

- Aggregate the data by time of day for aggregate volume in and out of the network
- Assess mean transit time of the network
- Assess key intersection flow and throughput
- Assess various traffic light timing sequences to enhance the traffic flow accordingly

The next section of this project report describes the different visualizations of the data that were produced for the sponsor to represent the data.

4 Visualization

The SUMO simulation produces an XML output file as shown in Figure 12 below. The XML file is then processes using a R program script to convert it to Comma Separated Value (CSV) format. The CSV file is then aggregated with other files from time and strata runs and the data is collected as shown in *Figure 15* below. That matrix is then sorted by entry/exit counts and relating to volumes of vehicles which are entered in the cells. They are then color coded with conditional formatting by visibility and displayed.

This display shows the number of vehicles exiting the defined area by location and time (blue (OUT) and the number of vehicles (Red (IN)).

Vehicle Count by Hour – Main Network

	5am	6am	7am	8am	9am	10am	11am	12pm	1pm	2pm	3pm	4pm	5pm	6pm	7pm	8pm
E10	34	52	32	40	61	23	42	42	23	23	21	48	40	31	40	62
E12	11	8	7	7	7	8	8	8	7	7	7	7	7	6	7	7
E26	9	15	12	15	15	3	15	15	6	6	3	12	15	9	15	15
E6	346	104	97	104	151	0	127	127	61	61	4	125	101	153	104	152
E7	195	79	161	97	140	45	85	85	125	125	97	106	118	80	97	135
H10	24	29	65	35	46	45	29	29	34	34	16	23	15	57	35	47
H11	8	5	10	8	10	6	9	9	2	2	5	6	7	9	8	6
H12	161	65	212	81	185	115	145	145	48	48	103	96	159	210	81	62
H4	232	176	233	189	157	124	189	189	125	125	57	114	221	230	189	155
H9	23	18	22	18	14	12	16	16	11	11	9	15	19	21	18	15
e13	23	20	22	22	4	0	22	22	8	8	0	10	24	25	22	4
e25	78	172	113	226	173	64	179	179	16	16	28	199	227	74	226	139
i24	15	22	26	21	9	12	20	20	0	0	12	24	26	18	21	9
i30	376	366	405	376	131	148	389	389	119	119	202	278	486	375	376	123

Figure 9: Number of Vehicles Entering and Exiting the Region of Interest by Hour

A summary of the overall aggregated data from the OUT and IN is collected and plotted on the line graph displayed in Figure 10 below. A similar process to that described is used for the aggregate the in and out data for the vehicles which is displayed in Figure 11 below. This second graph removes internal traffic that circulates in the area and does not qualify for the ASSETT, Inc. customer require of entry and exit of the defined area. The figure better aligns with the incremental flow of data.
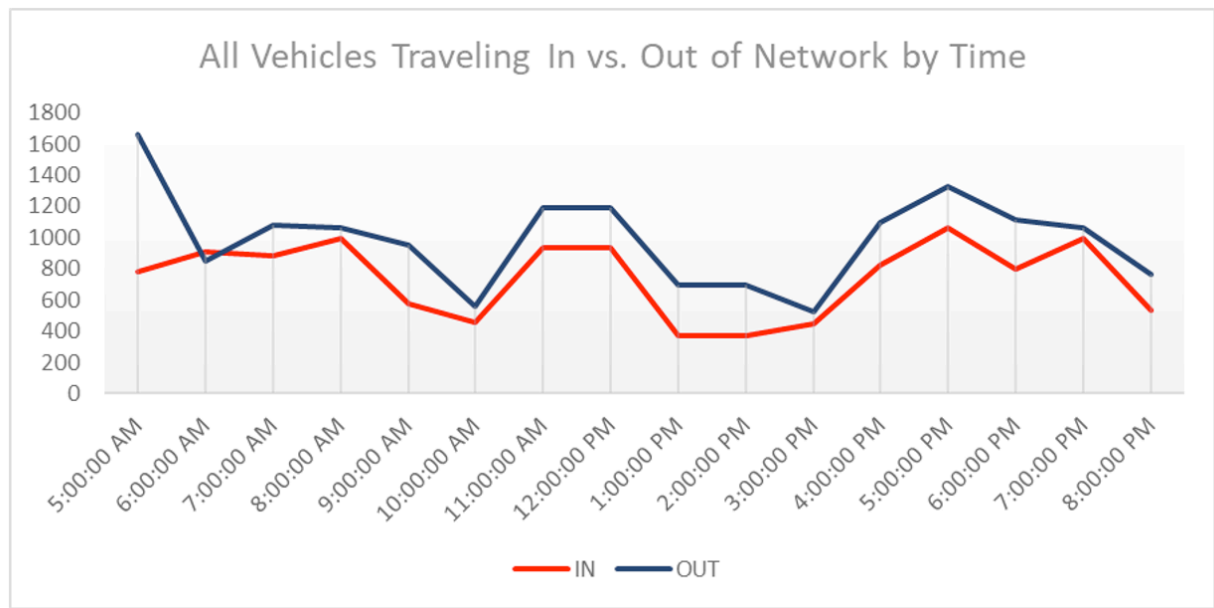

Figure 10: Total number of Vehicles Entering and Exiting the Region per Hour
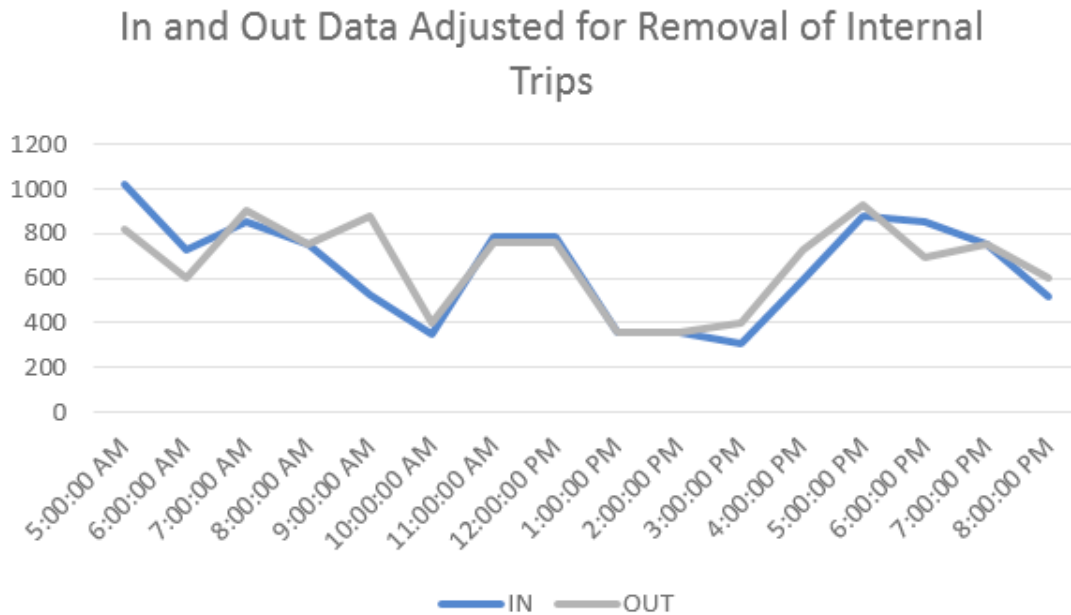
Figure 11: Graph of Median Transit Times of the Network by Hour

Other statistical methods could be applied to the data to assess the overall trends of the data such as a smoothing function to alleviate the dynamic changes in the data. These options were assessed but were not implemented due to the expectation that the customer wanted to see the greatest fidelity possible and a number or smoothing functions would manipulate the data from the raw data collected. A summary display is also provided for the entire quantity entering and exiting the area by hour.

Below is the raw XML file (*Figure 12)* produced by the SUMO simulation that is then parsed and converted to a CSV file for further processing. Additionally, shown is a snip of the CSV converted output file:

```
<additional xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="http://sumo.dlr.de/xsd/additional_
file.xsd">
    <routeProbe id="00" edge="-20420582#12" freq="3600"
name="neighborhood" file="Summary" begin="0.00"/>
    <routeProbe id="E1" edge="248292846#2" freq="3600"
file="routeProbe_248292846#2_0.xml" begin="0.00"/>
    <routeProbe id="E10" edge="-45781827#2" freq="3600"
name="Road 234" file="Summary" begin="0.00"/>
    <routeProbe id="E12" edge="687358734#3" freq="3600"
name="Road 234" file="Summary" begin="0.00"/>
    <routeProbe id="E14" edge="236147802#4" freq="3600"
name="Road 234" file="Summary" begin="0.00"/>
    <routeProbe id="E15" edge="237202564#6" freq="3600"
name="Road 234" file="Summary" begin="0.00"/>
    <routeProbe id="E19" edge="562353550#0" freq="3600"
name="Road 234" file="Summary" begin="0.00"/>
    <routeProbe id="E26" edge="229835698#3" freq="3600"
name="Intersection Road 234" file="Summary" begin="0.00"/>
    <routeProbe id="E27" edge="505036864#4" freq="3600"
name="Godwin Drive" file="Summary" begin="0.00"/>
    <routeProbe id="E28" edge="-145384937#2" freq="3600"
name="Godwin Drive" file="Summary" begin="0.00"/>
    <routeProbe id="E29" edge="276386083#2" freq="3600"
name="Godwin Drive" file="Summary" begin="0.00"/>
    <routeProbe id="E31" edge="217393307" freq="3600"
name="Godwin Drive" file="Summary" begin="0.00"/>
    <routeProbe id="E32" edge="-45823765#12" freq="3600"
name="Welington Rd" file="Summary" begin="0.00"/>
    <routeProbe id="E35" edge="652046740#0" freq="3600"
name="Wellington Rd" file="Summary" begin="0.00"/>
    <routeProbe id="E36" edge="686441311#0" freq="3600"
name="Ashton Ave" file="Summary" begin="0.00"/>
    <routeProbe id="E4" edge="562572902#6" freq="3600" name="Road
234" file="Summary" begin="0.00"/>
```

Attribute:time	edge.lane.vehic le.Attribute:id	edge.lane.vehic le.Attribute:pos	edge.lane.vehic le.Attribute:spe ed	edge.lane.Attri bute:id	edge.Attribute:i d	time
3	2060	5.1	21.33	20585577#2_0	20585577#2	5am
4	2060	25.49	20.39	20585577#2_0	20585577#2	5am
5	2060	46.11	20.62	20585577#2_0	20585577#2	5am
6	2060	66.69	20.58	20585577#2_0	20585577#2	5am
7	2060	86.89	20.2	20585577#2_0	20585577#2	5am
8	2060	108.1	21.21	20585577#2_0	20585577#2	5am
9	2060	128.72	20.63	20585577#2_0	20585577#2	5am
10	2060	150.04	21.32	20585577#2_0	20585577#2	5am
11	2060	171.28	21.24	20585577#2_0	20585577#2	5am
12	2060	191.64	20.35	20585577#2_0	20585577#2	5am
13	2060	212.53	20.89	20585577#2_0	20585577#2	5am
14	2060	233.19	20.66	20585577#2_0	20585577#2	5am
15	2060	253.75	20.56	20585577#2_0	20585577#2	5am
16	2060	274.89	21.14	20585577#2_0	20585577#2	5am
17	2060	295.12	20.23	20585577#2_0	20585577#2	5am

Figure 12: XML and CSV Output File from SUMO Simulation

The configuration files that are input into the SUMO configuration allows a tuning of traffic flow and volumes from-location to the destination to-location. The file below is representative of one of the configuration files used in the configuration of the system. The example file is shown in *Figure 13* below.

FROM	TO	COUNT	TYPE	FORMULA
E5	E17	33	THROUGH TRAFFIC	E5 E17 33
E16	E6	33	THROUGH TRAFFIC	E16 E6 33
E7	E2	33	THROUGH TRAFFIC	E7 E2 33
I30	E25	33	THROUGH TRAFFIC	I30 E25 33
I24	I30	33	THROUGH TRAFFIC	I24 I30 33
E26	E10	33	THROUGH TRAFFIC	E26 E10 33
E28	E4	33	THROUGH TRAFFIC	E28 E4 33
E31	E29	33	THROUGH TRAFFIC	E31 E29 33
E31	I24	33	THROUGH TRAFFIC	E31 I24 33
E12	H12	33	THROUGH TRAFFIC	E12 H12 33
E13	H11	33	THROUGH TRAFFIC	E13 H11 33
H18	E27	33	THROUGH TRAFFIC	H18 E27 33
H8	E8	33	THROUGH TRAFFIC	H8 E8 33
E7	E2	33	THROUGH TRAFFIC	E7 E2 33
E1	H9	33	THROUGH TRAFFIC	E1 H9 33
H10	routeProbe_0	33	THROUGH TRAFFIC	H10 routeProbe_0 33
S14	E35	33	LEAVING FROM INSIDE TO OUTSIDE	S14 E35 33
S13	H8	33	LEAVING FROM INSIDE TO OUTSIDE	S13 H8 33
S6	E8	33	LEAVING FROM INSIDE TO OUTSIDE	S6 E8 33
S4	E13	33	LEAVING FROM INSIDE TO OUTSIDE	S4 E13 33
S15	E4	33	LEAVING FROM INSIDE TO OUTSIDE	S15 E4 33
S3	E6	33	LEAVING FROM INSIDE TO OUTSIDE	S3 E6 33

Figure 13: Designated Route locations (From – To) by count and type for the Simulation

Figure 14: Map of Entry and Exit locations by ID located on the Area Map

The results of the SUMO analyses are aggregated into the format shown in Figure 15 below.

Vehicle Count by Hour - All Network																
Aston Ave & Balls Ford	34	52	32	40	61	23	42	42	23	23	21	48	40	31	40	62
Balls Ford & Sudley Rd	11	8	7	7	7	8	8	8	7	7	7	7	7	6	7	7
Aston Ave & Godwin Dr	9	15	12	15	15	3	15	15	6	6	3	12	15	9	15	15
Southbound 234/PrWillPkwy	346	104	97	104	151	0	127	127	61	61	4	125	101	153	104	152
Northbound 234/PrWillPkwy	195	79	161	97	140	45	85	85	125	125	97	106	118	80	97	135
University Blvd & Godwin Dr	24	29	65	35	46	45	29	29	34	34	16	23	15	57	35	47
North Godwin Drive & Sudley Rd	8	5	10	8	10	6	9	9	2	2	5	6	7	9	8	6
South Godwin Drive & Sudley Rd	161	65	212	81	185	115	145	145	48	48	103	96	159	210	81	62
Southbound Sudley Rd	82	43	85	43	35	28	52	52	59	59	17	53	73	83	43	40
Sudley Manor & Sudley Rd	190	94	92	134	110	54	187	187	152	152	59	187	318	58	134	49
Sudley Manor & 234/PrWillPkwy	232	176	233	189	157	124	189	189	125	125	57	114	221	230	189	155
Westboard University Blvd	225	240	213	264	124	147	241	241	164	164	150	230	241	222	264	123
Southbound University Blvd	23	18	22	18	14	12	16	16	11	11	9	15	19	21	18	15
Northbound Sudley Rd	23	20	22	22	4	0	22	22	8	8	0	10	24	25	22	4
Godwin Drive	78	172	113	226	173	64	179	179	16	16	28	199	227	74	226	139
Southwest Bound Godwin Drive	15	22	26	21	9	12	20	20	0	0	12	24	26	18	21	9
Northeast Bound Godwin Drive	376	366	405	376	131	148	389	389	119	119	202	278	486	375	376	123
Southbound Aston Ave	406	253	153	378	154	177	366	366	109	109	178	388	296	257	378	151
	5am	6am	7am	8am	9am	10am	11am	12pm	1pm	2pm	3pm	4pm	5pm	6pm	7pm	8pm

Figure 15: Matrix of Vehicles Transiting the Designated Area by Exit Location and by Hour

To ensure the alignment of our generated vehicle data with that of the customer collected survey data we identified specific points of correlation to ensure mapping of rough volumetric at key intersections to ensure we had alignment between the data sets. The analysis performed is summarized in *Figure 16* below:
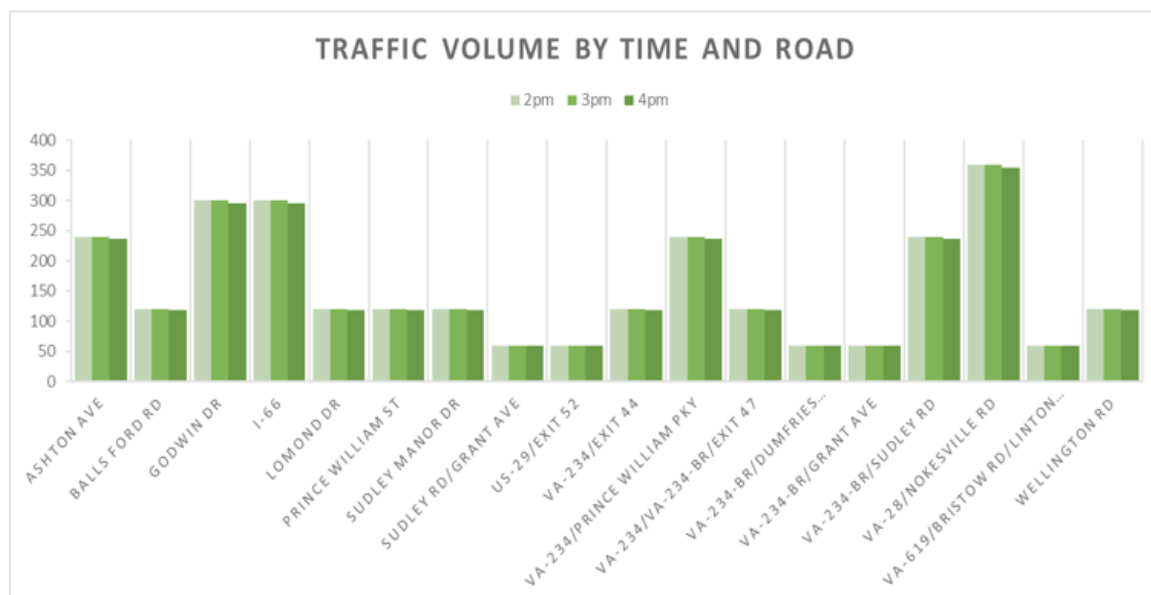

Figure 16: Summary Analysis for Correlation of Key Intersection Data

The customer data, while comprehensive, was only for a short period from 2:00pm-6:00pm on one day, and additional data was required to be collected to provide enough of a statistical sample. The HERE, Inc. data set was available for more data, but the purchase of data was beyond the scope or budget of the project. This may be an option for a follow-on effort but was not available to our team.

5 Findings

Our results show the volume of traffic by road for key areas within the network. The heat map allows our sponsor, ASSET, Inc., to identify the areas with the highest volumes of traffic (red color) and perform deeper analysis to identify possible solutions. As mentioned in the Future Work section of this document, there are a variety of steps one could take to further this analysis. This can include analyzing how adding a lane, incorporating emergency vehicles or pedestrians, or home development effect traffic flow on particular lanes and how this may affect commute times and community satisfaction.

Additionally, it was assessed that the median transit time of the main enter/exit vehicles was 21 minutes. This was calculated by taking the "end" time of the vehicles route (*Figure 10b -column "Attribute: time"* and subtracting the vehicles "begin" time. The vehicles that "hit" one of the main (entry/exit) "detectors" were included in the calculation, while any vehicles that only cycled within the network, were removed. While this is lower than the 39 minutes previous cited, this value only considers the vehicles transit within the network, and not their transit on nearby highways. A goal for further analysis would be to identify the routes with the highest median transit time and identify the solutions that might lower the median transit time without negatively impacting nearby roads.

The SUMO tool is a very powerful and multidimensional tool for analysis. The depth of the tool provides significant flexibility to assess a wide variety of conditions and characteristics. The extensions and options of the tool allow extensive analysis bit also introduce complexity and configuration challenges to work through. Our GMU ASSETT Team, analysis while successful, was hampered due to the sparse documentation of the tool and the interaction and interconnectedness of the various elements required to perform the simulation. The documentation is not clear for introductory users and the demonstration and tutorial information is sparse to learn such a tool.

We used the data analytic skills developed in the DAEN program for interpretation of data and tools necessary to accomplish this task. While it was an unstructured task, the customer and the faculty aided in scoping and focusing the effort to provide maximum value. The model and analyses have met and exceeded the customers' expectations (Mr. Jim Villani, ASSETT Inc, 11/15/2019)

Our analysis required us to assess and obtain significant knowledge in traffic engineering and analysis from open source resources on the web. We specifically used VDOT data for characterizing of traffic in the Prince William County area and the national statistics from the Census bureau for development of "profiles" of traffic. The great challenge is in assessing "routes" for vehicles. While we have data on specific locations and time of traffic volumes the source and destination of traffic is very helpful in determining flow and suggesting alternate routes or alternatives in general for travelers as options to reduce time and congestion. This data was imputed from population characteristics from VDOT, US Census, Northern Virginia local websites, and Prince William County data. [10]

6 Summary

In summary the GMU ASSETT team was able to fulfill our customer's need and provide a specific data model for traffic flow volume, and delay in the Manassas, Virginia region provided. We also developed a tool that the customer could use to further refine and assess parameters of interest. While we did not use

the HERE, Inc. data due to cost and storage issues, that is an option is available for our customer to pursue based on their business priorities. A series of scripts could be developed to ingest the HERE data and use it as input to the FLOW model and is proposed as a follow-on task for a future cohort in the GMU DAEN 690 program.

- The SUMO tool produced data that correlates to the collected customer data
- The sumo tool allows a detailed analysis and tuning of the data to assess impacts and alternatives
- The data visualization enables traffic engineering and traffic management personnel to develop increasingly complex and interoperable models for traffic engineering, emissions control, road improvement options, and many other scenarios

Our work is only a first step in a variety of analytic application for future looking analysis that could be provided for using artificial intelligence and deep learning, reinforcement Learning, in Northern Virginia Traffic engineering and analyses. The tools and techniques could vastly improve the top challenges to our region which are: traffic congestion, carbon emissions, and lost time due to traffic incidents.

7 Future Work DAEN 690 - Follow on Research in Traffic analysis and Simulation.

Introduction: As a result of the initial study conducted in Fall of 2019 for DAEN 690 at George Mason University a follow-on deep learning task has been identified for a subsequent team to address in a semester long project session. The initial project was to assess traffic congestion in the Manassas area and to assess traffic volumes within a specific bounded area. The Secondary objective was to reduce median transit time for divers in the area. A full project report detailing the project is available from GMU DAEN 690 professors.

Key Terms: Data analytics, Deep Reinforcement Learning, Traffic Modeling and Simulation, SUMO

Concept for the Simulation of Urban Mobility (SUMO)

"Simulation of Urban MObility" (SUMO) is an open source, highly portable, microscopic road traffic simulation package designed to handle large road networks. It is mainly developed by employees of the Institute of Transportation Systems at the German Aerospace Center. SUMO is licensed under the EPL. [3]

The concept for this project is to enhance the Fall 2019 DAEN 690 traffic simulation developed in SUMO, NetEdit, and TraCI with an artificial deep reinforcement learning model. This model will extend the manual interpretation of traffic patterns and flows. The proposed follow on activity would be to incorporate deep reinforcement learning with the FLOW module into the architecture and enable reinforcement learning to adjust and adapt the volumetric data and flow directions to determine optimal parameters for a given network.

During the fall 2019 sessions there was not enough time to assess the problem identify the advanced technology systems, trade of the systems develops a model and enhance the model for optimal performance. It was not until approximately ½ way through the semester that the team learned to the FLOW option and the Reinforcement Learning opportunity.

Therefore, the enhanced traffic analysis concept for the follow-on effort proposes a computational framework and architecture to systematically integrate deep Reinforcement Learning (RL) and traffic microsimulation, thereby enabling the systematic study of autonomous vehicles in complex traffic settings, including mixed-autonomy and fully autonomous settings. The SUMO and FLOW framework permits both RL and classical control techniques to be applied to microsimulations. As classical control is a primary approach for studying traffic dynamics, supporting benchmarking with such methods is crucial for measuring progress of learned controllers. The computational framework encompasses model-free reinforcement learning approaches, which complement model-based methods such as model-based reinforcement learning, dynamic programming, optimal control, and hand-designed controllers; these methods dramatically range in complexity, sometimes exhibiting prohibitive computational costs.

The contribution of DAEN690 and other referenced literature includes three components, (1) a computational framework and architecture, which provides a rich design space for traffic control problems and exposes model-free RL methods, (2) the implementation of several instantiations of RL algorithms that can solve complex control tasks, and (3) a set of use cases that illustrates the power of the building block and benchmark scenarios.[1].

Flow, a computational framework for deep RL and control experiments for traffic microsimulation. Flow integrates the traffic micro-simulator SUMO with a standard deep reinforcement learning library rllab, thereby permitting the training of large-scale reinforcement learning experiments at scale on Amazon Web Services (AWS) Elastic Compute Cloud (EC2) for traffic control tasks. The computational framework is open-source and available at https://github.com/cathywu/flow.

Implementation

The Fall 2019 DAEN 690 team used a windows platform and integrated modules, after further analyses the Linux platform on AWS would better serve the architecture and would provide further scalability and flexibility. At the start of the semester these resources were not available but have been incorporated into the program providing enhanced resources and capabilities for more advanced projects. The SUMO software is available on the Internet and is freely downloadable with enhanced features and modules https://sumo.dlr.de/docs/SUMO_User_Documentation.html

References

 [1] Wu, C., Kreidieh A., Et. al Flow: Architecture and Benchmarking for Reinforcement Learning in Traffic Control https://www.arxiv-vanity.com/papers/1710.05465/

[2] Sang Koh S. Zhou B, Et al. (2018) Reinforcement Learning for Vehicle Route Optimization in SUMO IEEE Xplore Digital Library. https://ieeexplore.ieee.org/abstract/document/8622980

[3] SUMO Documentation https://sumo.dlr.de/docs/SUMO_User_Documentation.html

Appendix A - Code references – any code used for the analysis

Appendix B- Risk Section

Appendix C- Agile Development

Appendix D - References

[1] Khodakaram Salimifard and Mehdi Ansari (2013). Modeling and Simulation of Urban Traffic Signals International Journal of Modeling and Optimization, Vol. 3, No. 2, April 2013 https://nacto.org/docs/usdg/modeling_and_simulation_of_urban_traffic_signals_salimifard.pdf

[2] Suni Pegu, Thuleswar Nath. (2017) A review on modelling and simulation of vehicular traffic at signalized intersection A review on modelling and simulation of vehicular traffic at signalized intersection International Journal of Engineering Trends and Technology (IJETT) – Volume 48 Number 5 June 2017 https://pdfs.semanticscholar.org/ba37/27cad71e70d61b3bb6d405a4e75c6831f84f.pdf

[3] Flow Project https://flow-project.github.io/

[4] Eclipse SUMO – Simulation of Urban Mobility https://www.dlr.de/ts/en/desktopdefault.aspx/tabid-9883/16931_read-41000/

[5] The Transportation Analysis and Simulation System (TRANSIMS) https://sourceforge.net/projects/transims/

[6] PVT Visum http://vision-traffic.ptvgroup.com/en-us/products/ptv-visum/

[7] The ONE simulator https://akeranen.github.io/the-one

[8] John Hourdakis, Panos G. Michalopoulos, Jiji Kottommannil (2003). Practical Procedure for Calibrating Microscopic Traffic Simulation Models Transportation Research Record, vol. 1852, 1: pp. 130-139., First Published Jan 1, 2003. https://journals.sagepub.com/action/doSearch?target=default&ContribAuthorStored=Kottommannil%2C+Jiji

[9] Daniel Krajzewicz, Jakob Erdmann, Michael Behrisch, and Laura Bieker. (2012). Recent Development and Applications of SUMO - Simulation of Urban MObility. International Journal on Advances in Systems and Measurements, 5 (3&4):128-138, December 2012.

[10] (2016) Virginia Department of Transportation. Vehicle type by highway in 2016. http://www.virginiadot.org/info/resources/Traffic_2016/AADT_PrimaryInterstate_2016.pdf

[11] (2019). index mundi:https://www.indexmundi.com/facts/united-states/quick-facts/montana/average-commute-time#map

[12] Vandyck T., Rutherfors T. (10, 2018). Regional labor markets, commuting, and the economic impact of road pricing. Elsevier Regional Science and Urban Economics Volume 73, November 2018, Pages 217-236.

[13] Waddell j. (2014) Are Long Commutes Killing Your Employee Productivity? Business 2 Community, Business & Finance, Human Resources https://www.business2community.com/human-resources/long-commutes-killing-employee-productivity-0848737

[14] Paton N. (2017) Long commutes to work take toll on employee health Occupational health & Wellbeing https://www.personneltoday.com/hr/long-commutes-work-take-toll-employee-health/

[15] Bruno J. (2019). ASSETT Inc provided data from Here Inc extracts of data