CMSC 691 DATA PRIVACY PROJECT REPORT

TOPIC – Location Privacy- K Anonymity & Obfuscation (Food Delivery System)

Sai Karthik Dattu **Saikrishna Dugyala** **Sai Sandeep Ravuri**

WZ30691 GU24496 YL30934

ABSTRACT:

In this project, we have created a novel technique to establish location privacy based on both K Anonymity and Obfuscation. By concentrating more on user preferences and privacy, we have developed ways of safeguarding location privacy. Our technology enables users to share their location with delivery personnel while ensuring that all operations are secure, and that location privacy is maintained according to the user's choices. The delivery person will never know the specific location information of users because K anonymity with obfuscation is implemented, and most of the time a new location point is generated, providing enough randomization for location privacy, and making the solution an optimal one. This approach also allows customers to choose how far they are willing to travel.

INTRODUCTION:

Nowadays, most user services, such as food applications, mail delivery systems, and COVID tracking systems, are based on location data. Due to various current technological improvements, maintaining the user's location privacy has become one of the most difficult issues. Due to faulty data collection techniques, unlawful selling of user personal data, untrustworthy service providers, and third-party applications, maintaining location privacy has been difficult.

Many solutions aimed at preserving location data by focusing on avoiding user location data leaks by encrypting location data in transit and at rest, and by constructing secure channels while transferring location data.

In our study, we address user privacy concerns connected to location data and have proposed solutions by providing varying degrees of privacy based on user preferences.

MOTIVATION:

The main motivation behind our project is, recently we have come across an incident where we ordered some food from a particular restaurant to be delivered to our location. And saw that the person who is going to deliver our order, previously delivered few orders to our location. From this we can say that this person knows our location, from which restaurant we are ordering food and what type of food we order. This is a simple case of privacy not being protected. But there are much more serious cases like some of the telecom companies selling the real-time location of phones to shady companies which is reported by VICE news, apps that track location data may turn around and sell that data, revealing someone's every movement, a company with thousands of cameras selling car locations to debt collectors and others, The worst part is in many of the states covering the car's license plate is illegal. The debate is particularly difficult when you look at the successful use of location data during the covid times across different governments. Systems can track the movements of people who are sick and order those likely exposed to Covid-19 to self-quarantine. Public health authorities could use the data to identify hot spots, allocate scarce resources to hospitals, and give advance warnings of larger outbreaks to state and federal authorities. All of this would come at an enormous cost to our privacy. The risks of location tracking outweigh the benefits. So, we tried to minimize the risk and balance it with the benefits. Our main aim is to strike a balance between privacy and utility to give an optimum result to the user.

PRIVACY MODEL:

In this model we have taken an example of location-based service which is food delivery system and conducted an experiment on this. In this model every user in the system has this food delivery app and by using this user can place an order from his favorite restaurant. While placing the order user needs to give his location data because the restaurants and the delivery person can use it to deliver the food. We consider both the restaurants and the delivery person to be untrustworthy in this circumstance, and if the user submits the request directly to the restaurants, his or her name may be revealed, compromising his or her privacy. Motivated by this fact we have developed our system in such a way that when user places the order it will not directly submit the location data of user to restaurants instead it asks the users to choose the degree of anonymity he wants and after anonymizing the user location data we will obfuscate it to provide enough randomization for the user’s data. In addition to this mechanism, we have also offered another solution to user based on his preferences like the amount of distance he’s willing to travel. We have also removed any other obvious identifiers which will disclose the user’s information.

ASSUMPTIONS AND THREAT MODEL:

For the provision of privacy in LBSs, a few centralized K–anonymity approaches have been proposed. The working assumptions about the attacker's capabilities that are used by most of these approaches are presented in the sections that follow. The assumptions must be known to compare the various techniques in terms of the privacy guarantees they provide to LBS requesters. The following abilities are assumed to be present in the attacker:

1. An attacker can intercept the region where the requester of an LBS is offered anonymity. This indicates that the LBS provider is not trustworthy.
2. The attacker is aware of the methods that the trusted server employs to provide privacy in LBSs. This is a regular occurrence in the security literature, where algorithms are frequently made public.
3. The attacker has access to all the system's users' present locations. The fact that users frequently send searches from readily recognizable places supports this idea. Because it's impossible to predict the exact quantity of information an attacker has, this assumption necessitates that the privacy mechanism be proven secure in the worst–case situation.
4. The attacker tries to breach the location privacy of the users by using only current location data; he/she is unaware of any historical information about the movement of the users, as well as any behavior patterns of clients (e.g., a user is often asking a particular query at a certain location or time).

METHODOLOGY:

There are basically two approaches for K-anonymity, which is data dependent and Space based cloaking, we have chosen data dependent cloaking over other approach because due to the limited time and resources we have, we were not able to generate the grid models for the location. Even though we found out few resources that can generate the grid models, they were mostly third-party applications and not available freely. Contrast to this we found resources like GCP, Maps SDK by google to implement the data dependent cloaking.

In data dependent cloaking there are two methods k-bucket cloaking and distance-based cloaking.

K-bucket Cloaking:

In k we are going to give a certain value, in this case if it is 4, we are going to find at least k-1 houses and generate a minimum bounding circle. The radius of the circle is determined by k-1 houses, for suppose this is where our actual location is, and this is the farthest house from the center. We are going to take the distance between the center and the farthest house in k-1 houses as the radius and draw a minimum bounding circle. To extend this algorithm, we have used this circle as an input to the obfuscation algorithm that we have implemented. This algorithm randomly generates different set of latitudes and longitudes in the circle. And we are checking the integrity of this location to see that it is not at the middle of a lake or anything like that. Also, for this randomly generated location we are going to find the nearest road so that the order can be delivered. We have done all this validation and finding nearest road using the Maps SDK by Google and GCP.

Distance -based Cloaking:

To make this solution more user friendly, we have introduced distance-based cloaking, in which the user specifies a distance ‘d’ and generate a circle C1. A center for all the houses which are in the circle C1 is found generating a new circle C2 with the radius as distance between center point and user’s exact location, a random point is generated in this circle.

Conclusions and Future References:

When we performed our experiment in the community with very less population (communities like Berkley springs in west Virginia and silver springs community in Maryland), we have lost good amount of utility as the population is very less and houses count is few. But when we performed same experiment in the area with high density population, the utility loss is very less as the houses are near to each other. From the above experiments we can say that utility loss is high areas with less houses and utility loss is less in area with more houses. For the future improvements we are planning to suggest user the K and D values based on the user county or community without accessing the user’s exact location, with this we are preserving user’s location.

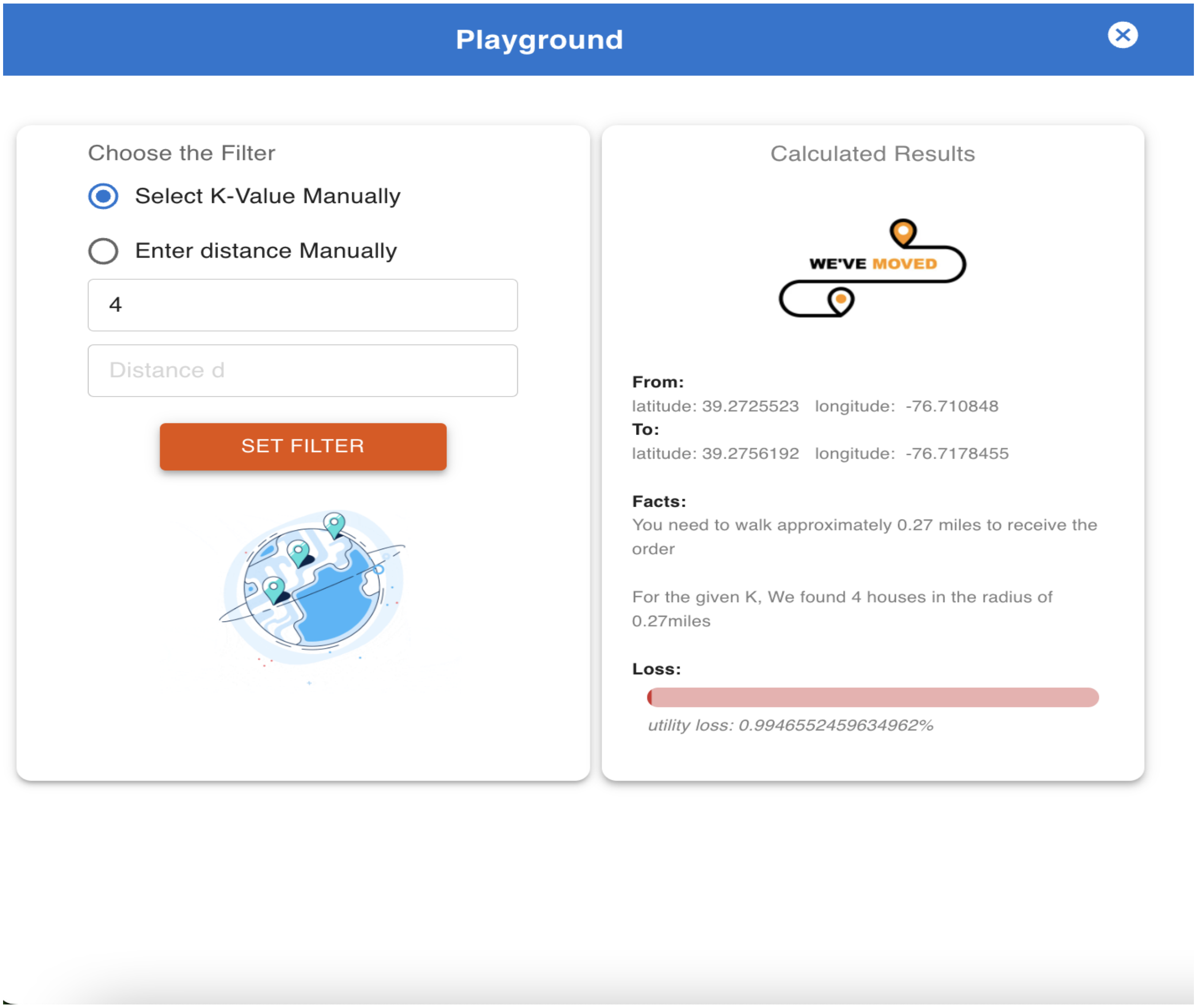
Individual Contributions:

Sai Karthik Dattu: Designed APIs to fetch nearest houses in the given radius and designed API to find nearest road. Worked on GCP and frontend designing

Sai Sandeep Ravuri: Introduced Distance based cloaking with obfuscation in the paper and integrating the algorithm in the project.

Sai Krishna Dugyala: Introduced K-bucket cloaking with obfuscation in the paper and designed the algorithm in the project.

Sample images of the project:



Choose the Filter

● Select K-Value Manually

○ Enter distance Manually

4

Distance d

Calculated Results

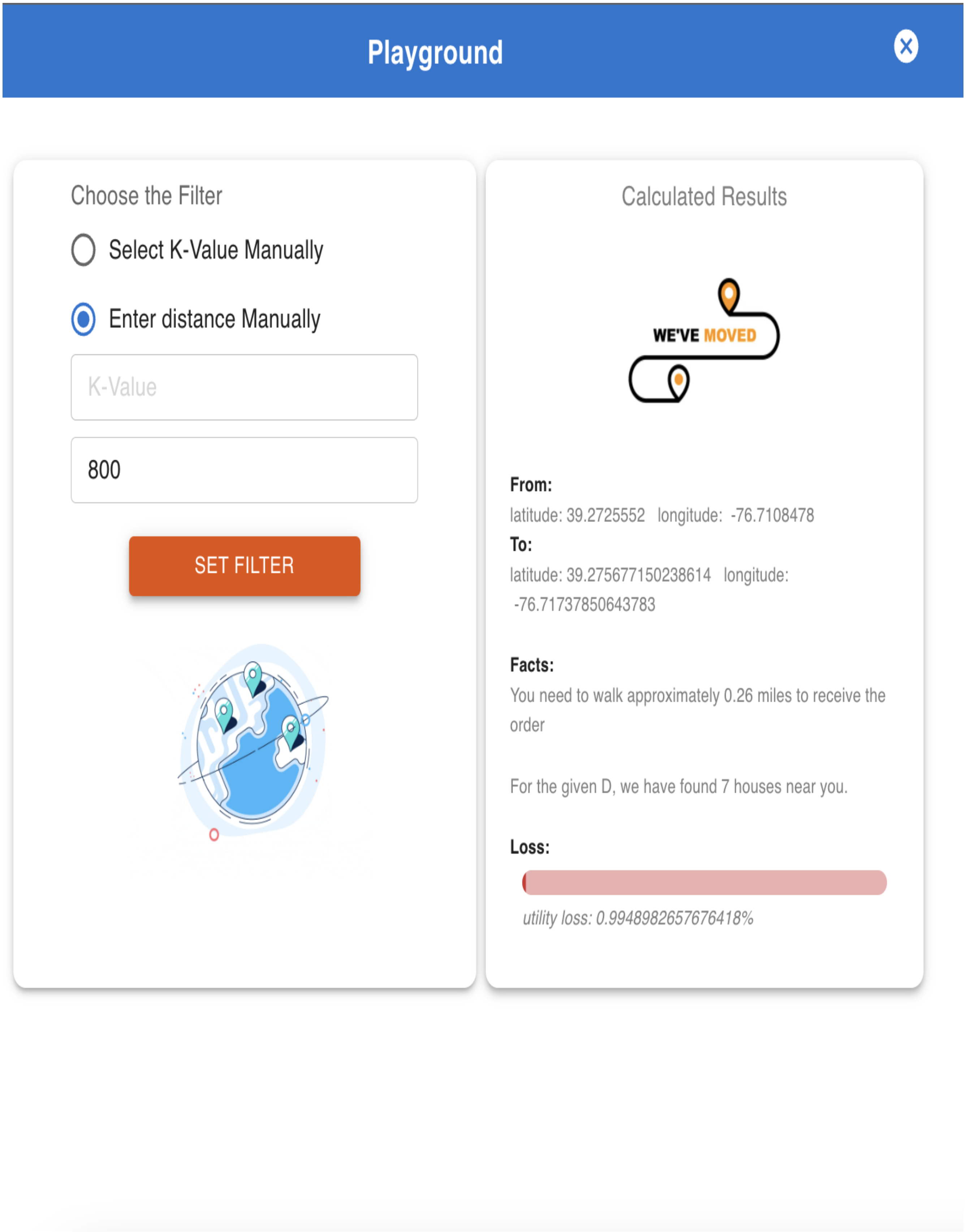
From:
latitude: 39.2725523 longitude: -76.710848
To:
latitude: 39.2756192 longitude: -76.7178455

Facts:
You need to walk approximately 0.27 miles to receive the order

For the given K, We found 4 houses in the radius of 0.27miles

Loss:

utility loss: 0.9946552459634962%



Playground

Choose the Filter

○ Select K-Value Manually

● Enter distance Manually

K-Value

800

SET FILTER

Calculated Results

From:

latitude: 39.2725552 longitude: -76.7108478

To:

latitude: 39.275677150238614 longitude: -76.71737850643783

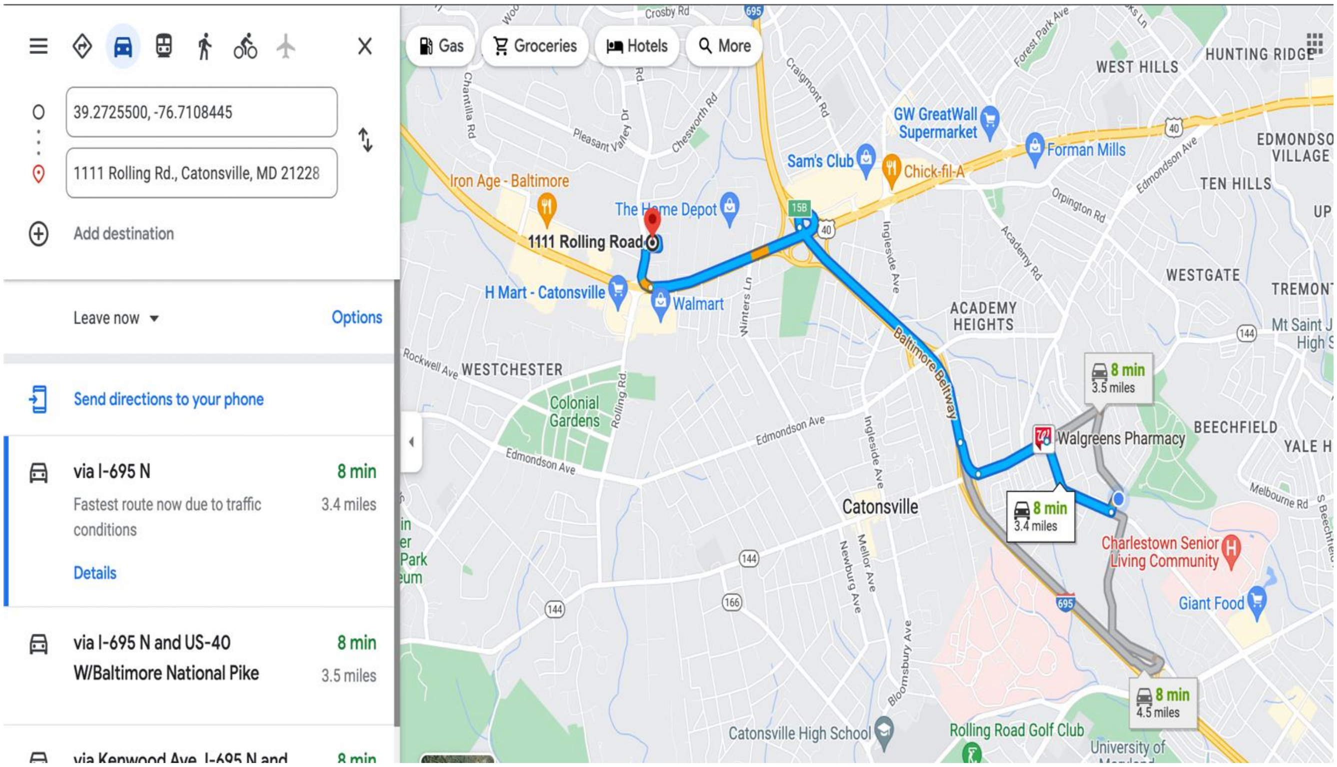
Facts:

You need to walk approximately 0.26 miles to receive the order

For the given D, we have found 7 houses near you.

Loss:

utility loss: 0.9948982657676418%



39.2725500, -76.7108445

1111 Rolling Rd., Catonsville, MD 21228

Add destination

Leave now ▾ Options

Send directions to your phone

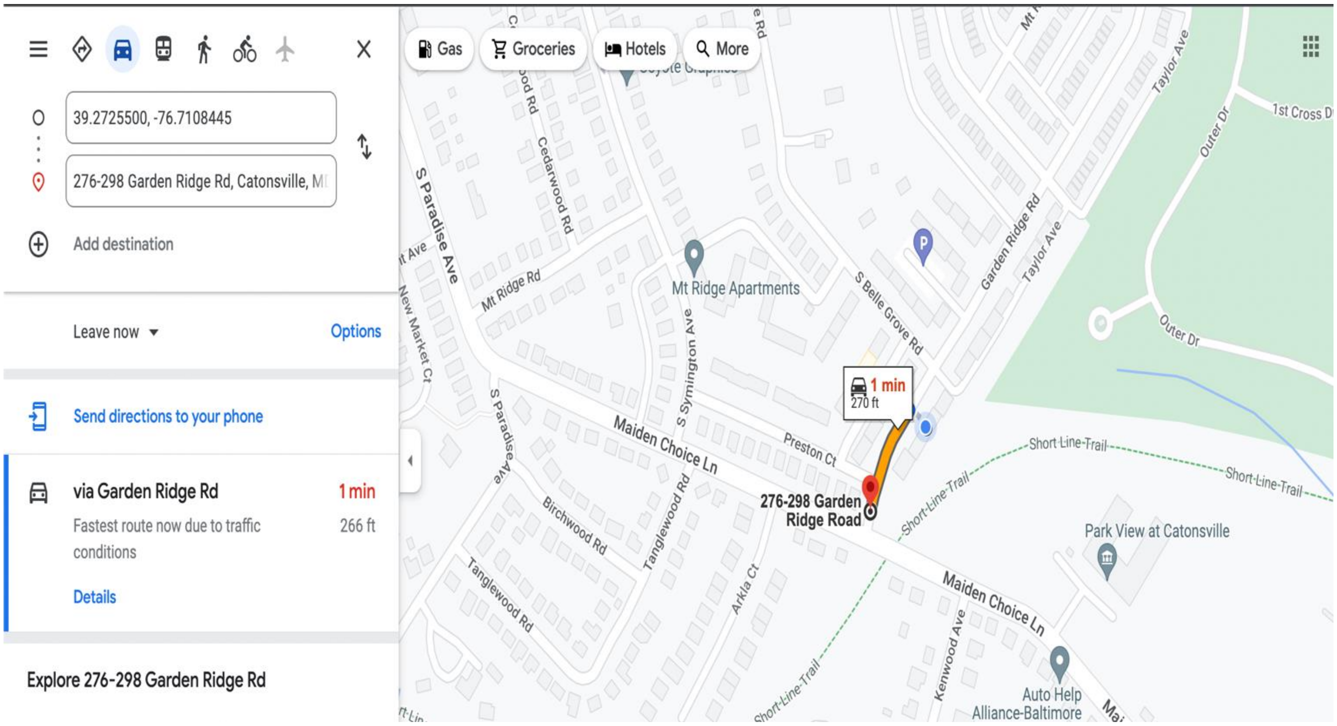
via I-695 N — 8 min — 3.4 miles

Fastest route now due to traffic conditions

Details

via I-695 N and US-40 W/Baltimore National Pike — 8 min — 3.5 miles

via Kenwood Ave, I-695 N and — 8 min



39.2725500, -76.7108445

276-298 Garden Ridge Rd, Catonsville, M

Add destination

Leave now ▾ Options

Send directions to your phone

via Garden Ridge Rd — 1 min — 266 ft

Fastest route now due to traffic conditions

Details

Explore 276-298 Garden Ridge Rd

References:

1. https://www.kdd.org/exploration_files/v12-1-p3-gkoulalas-sigkdd.pdf

2. https://www.youtube.com/watch?v=-j-H7U0I6Ao&t=973s

3. https://simon-oya.github.io/files/oya-2017-11-ccs-slides.pdf

4. https://spdp.di.unimi.it/papers/pcac07.pdf