WGUPS Delivery Distribution System Overview

Stated Problem:

The Western Governor's University Parcel System needed to find efficient routes to distribute their packages in a time and distance efficient manner. Assumptions were as following: packages must be delivered by their deadlines, there are three trucks, two truck drivers, each truck travels at a constant 18 miles per hour, and an average of 40 packages are to be delivered each day. The problem closely mimics the traveling salesman problem, which is an NP-hard problem. The nearest neighbor algorithm worked well for this problem, and I decided to use it as the core algorithm in this program. The nearest neighbor algorithm is self-adjusting in the fact that different input results in different results. For all locations that each truck would travel to, the algorithm would use a current location and find the closest location to that current location. After the algorithm finds the closest location, it would add that location to a list that holds the route path. Then, for the next iteration, that closest location would be the current location, and the process repeats. With the given package data, the three trucks delivered all their packages in 105.2 miles.

Nearest Neighbor Algorithm Overview:

- 1. The algorithm takes in a graph, starting location, and a truck as input.
- 2. An empty list that holds the order to visit is created, and an unvisited list, which holds the locations in it that have not yet been visited in the algorithm.
- 3. While the unvisited list is not empty, the nearest neighbor algorithm iterates. If the loop is on the first iteration, it starts at the start vertex.
- 4. For every location that has to be visited, the distance between the current vertex is compared. The location that has the shortest distance is saved as the closest location.
- 5. All packages on the truck that have the address of the closest location are marked as delivered, and their time delivered timestamp is updated.
- 6. The order to visit list appends the closest location. The unvisited list removes the closest location.
- 7. The closest location is assigned to current location. Steps 3-6 are repeated until the length of the unvisited list is 1.
- 8. When the unvisited list is of length 1, the path string is created using the order to visit list. The algorithm returns the results, the order to visit, distance traveled, and the path string.

Nearest Neighbor Space-Time complexity in Big O Notation – $O(N^2)$

Nearest Neighbor Algorithm Pseudocode:

A. The algorithm takes in a graph that has all the locations to visit and distances between them, a starting location (vertex on the graph), and the truck in which nearest neighbor is finding an efficient route for.

B. Create order to visit, unvisited list, and distance traveled variable.

Set order to visit as an empty list

```
Set unvisited_list as an empty list
Set distance traveled to 0
```

C. Add all locations in the graph to unvisited_list and set the vertex class data member visited to false for each vertex so that the algorithm will know to visit each vertex.

```
for vertex in graph:
    append vertex to unvisited_list
    set vertex.visited to False

current vertex = starting location
order to visit append current vertex
```

D. Following code corrects package 9 information in the algorithm if package 9 is in the truck given as an argument.

```
for package in the truck.package_list:

If package.package_notes is 'Wrong address listed' visit_address is False

for packages in truck.package_list
    if packages.address equals package.address
    visit_address is True

else
    visit_address is False

package.visited is True

if visit_address is False
    for location in the unvisited_list
    if the location equals package address
    location.visited equals True

package_nine_corrected_time = 10:20
```

E. This while loop will find the nearest neighboring location, provide package updates, keep track of distance, and append the closest location to the order_to_visit list

```
while the length of unvisited_list > 0
    i = 1

if the length of unvisited_list is 1

if the truck has to return to hub
    closest location distance = distance between current vertex and start vertex
    distance traveled = distance traveled + closest location distance
    order_to_visit append start location (WGUPS)
    update truck's current time based on distance traveled

ff truck is 3 and truck's time is after package nine corrected time
```

update package 9's address and mark the package as visited and update status

```
get distance from current vertex to package 9
       update total distance traveled
       append package 9's address to order to visit
       update truck's current time based on distance traveled
     create a path string that shows the route of the order to visit list
     return order to visit, distance traveled, and path string
   for all vertexes in the graph's adjacency list of current vertex
     if i is 1 and vertex has not been visited
        closest location = vertex
        closest location distance = distance between start vertex and vertex
        i = i + 1
     elif i does not equal 1 and vertex has not been visited
       if the distance between the current vertex and vertex is less than the distance of closest
location distance
          closest location = vertex
          closest location distance = distance between current vertex and vertex
          i = i + 1
  for all packages
     if package address equals the closest location
       update package data members (time delivered, status, package.visited = True)
  order to visit append closest location
  mark the current vertex as visited so algorithm doesn't go back to current vertex
  Update truck's current time and also update distance traveled
```

B2: DEVELOPMENT ENVIRONMENT

The program was built using python version 3.9 and was developed on a MacBook Pro 13- inch, M1, 2020 on the macOS Big Sur version 11.3 operating system. I used the IDE, **PyCharm**, to build the program, and the version number of PyCharm used was 2021.1.2.

Space-Time and BIG-O:

| data_loader.py | Space Complexity | Time Complexity |
|----------------|-------------------------|---------------------|
| 61 | O(N) | O(N) |
| 90 | O(N^2) | O(N^2) |
| 134 | O(N^2) | O(N^2) |
| Total: | $N + 2N^2 = O(N^2)$ | $N + 2N^2 = O(N^2)$ |

| packages.py | Space Complexity | Time Complexity |
|-------------|--|--|
| 28 | O(1) | O(1) |
| 33 | O(1) | O(1) |
| 47 | O(1) | O(1) |
| 73 | O(1) | O(1) |
| Total: | $ \begin{array}{c} 1 + 1 + 1 + 1 = \\ O(1) \end{array} $ | $ \begin{array}{c} 1 + 1 + 1 + 1 = \\ O(1) \end{array} $ |

| truck.py | Space Complexity | Time Complexity |
|----------|-------------------------|----------------------|
| 13 | O(N^2) | O(N^2) |
| 259 | O(N^2) | O(N^2) |
| 289 | O(1) | O(1) |
| Total: | $1 + 2N^2 = O(N^2)$ | 1 + 2N^2 = O(N^2) |

| nearest_neighbor.py | Space Complexity | Time Complexity |
|---------------------|-------------------------|-----------------|
| 18 | O(N^2) | O(N^2) |
| Total: | O(N^2) | O(N^2) |

| user_interface.py | Space Complexity | Time Complexity |
|-------------------|-------------------------|-----------------|
| 13 | O(N^2) | O(N^2) |
| 85 | O(N^2) | O(N^2) |
| 212 | O(N^2) | O(N^2) |
| Total: | $3N^2 = O(N^2)$ | $3N^2 = O(N^2)$ |

| hash_table_chaining.py | Space Complexity | Time Complexity |
|------------------------|-------------------------|-----------------|
| 8 | O(N) | O(N) |
| 26 | O(N) | O(N) |
| 41 | O(N) | O(N) |
| 51 | O(N) | O(N) |
| Total: | 4N = O(N) | 4N = O(N) |

| main.py | Space Complexity | Time Complexity |
|---------|-------------------------|--------------------|
| 13 | O(N^3) | O(N^3) |
| 167 | O(1) | O(1) |
| Total: | $1 + N^3 = O(N^3)$ | $1 + N^3 = O(N^3)$ |

The Space and Time Complexities of the entire program are both, in big O notation, O(N³).

B4: Adaptability:

The program's current state was designed to handle three trucks, two truck drivers, and a maximum of 48 packages as stated in the requirements. The core algorithms in this program will need to be touched up if this program were to be scaled up. However, with minor changes, the program can be scaled up quickly. Some things would need to be known before scaling the program, like the number of trucks, the packages with wrong addresses and corrected times, and the range of the number of packages in which the program should handle. It is currently optimized for the requirements but can be optimized for other requirements as well. A concern I may have as the number of packages increase in the program would be the time efficiency. If the program were to handle thousands of packages, it would slow down, and some improvements to the algorithms would have to be made. The self-adjusting algorithm, nearest neighbor, would need small changes but can definitely handle a larger amount of data. The package data is stored in a hash table. Locations are stored in a graph as vertexes, and their distances apart from each other are stored in dictionaries. Both the hash and graph data structures are great for scalability. If new requirements were required of the program, I am sure that the program would work as intended with minor changes.

B5: Software Efficiency and Maintainability:

The Big-O time complexity of this program is $O(N^3)$. The program's Big-O time complexity is efficient but not considered fast; however, with the given input size, the program runs well. The program reads all the package, location, and distance data from the three CSV files, and if these files were to be updated, the package distribution on the trucks and the routes would change. The program is adaptable to new data for the majority of circumstances. If a spreadsheet were to be created with incorrect addresses and when those package addresses would

be updated, the program could have been built better for maintainability. Currently, only package nine's address must be updated at a specific time, and the program would not be able to update the address of a new package that has the wrong address listed. An algorithm in truck.py called truck_loader loads the packages onto truck that allows the trucks to make fewer stops. For new package data, the algorithm would need another update; packages that must be delivered together were manually added to a truck, and the algorithm would have to be updated to deliver packages together.

For future changes in the program, another developer would understand the logic in the code. All blocks of code have detailed comments explaining the logic behind the code. Also, functions are located in the files in which they would be expected to be found.

B6: Self-Adjusting Data Structures:

Throughout the program, there are two self-adjusting data structures, a graph and a hash table. Some strengths of the hash table for package data would be that it is efficient in inserting, searching, and removing, but a drawback would be that once the size of the hash table is set it cannot be resized. Also, I would not say I liked looping through the hash table for each package as it was not straightforward like a list would be and required extra steps. All the package data in the packages.csv file is inserted into the hash table every time the program is run and when a new package is inserted. Package data is updated frequently through the program, and it is easy to search for a package by just using the package id. The graph data structure holds all the locations from the package hash table. The graph data structure has two dictionaries, one for adjacent locations and another for the distances between locations. I chose the graph data structure because I liked that it was easy to visualize the data and that it worked well in implementing into the nearest neighbor algorithm. However, using a separate data structure for the locations required an extra step in the nearest neighbor algorithm to mark the package delivered. The algorithm had to loop through all packages in the hash table and look for packages with the address that was visited in the algorithm; if a package had that address, it was delivered, and some data members were updated.

D1, E, and F: Explanation of Hash Table:

The hash table is loaded with package data in the load_packages function in data_loader.py. The function retrieves the package data in packages.csv and creates a package object for each package that includes a package's id, address, city, state, zip, deadline, mass, and notes. The package is then inserted into the hash table using the package id as a key, the first argument, and the second argument is the package object created. Packages can be retrieved by searching for the package id. When a package is searched for and found, all package components and their values are displayed. A hash table is more efficient than a linear search because the hash function calculates the index of the item you are looking for, which is the package id in this program. The hash table will then look through all key and values in the index because the hash table I used for this program is a chain hash table. If the key equals the argument key, the value is returned. If only 1 item is inserted into each bucket, which happens to be the case for this program, the hash, insert, update, and remove perform at constant time complexity, O(1). On the other hand, a list's time complexity in big O notation is O(N).

- G. Provide an interface for the user to view the status and info (as listed in part F) of any package at any time, and the total mileage traveled by all trucks. (The delivery status should report the package as at the hub, en route, or delivered. Delivery status must include the time.)
 - 1. Provide screenshots to show the status of *all* packages at a time between 8:35 a.m. and 9:25 a.m.

The following screenshot shows the status of all packages at 8:45 a.m.

```
| Column | Section | Secti
```

2. Provide screenshots to show the status of *all* packages at a time between 9:35 a.m. and 10:25 a.m.

The following screenshot shows the status of all packages at 9:45 a.m.

```
Section 1. Section 1.
```

3. Provide screenshots to show the status of *all* packages at a time between 12:03 p.m. and 1:12 p.m.

The following screenshot shows the status of all packages at 12:15 p.m.

```
There for the leaves and second for the first of the firs
```

H. Screenshots of Code execution:

When the program is run:

```
/Users/dylanstahl/PycharmProjects/deliveryTruckRoutes/venv/bin/python /Users/dylanstahl/PycharmProjects/deliveryTruckRoutes/main.py
Welcome to Western Governors University Parcel Service (WGUPS), this program has found an efficient route and delivery distribution for Daily Local Deliveries (DLD).

Press 1 to check the status of a package
Press 2 to view the status and info of all packages at a specified time, and the total mileage driven by each truck
Press 3 to view the end of day result
Press 4 to insert a new package
Press 5 to reset the packages to the original list
Type 'quit' to exit the application
```

When the end of day result is printed (Total miles driven is shown at the bottom):

```
Tool 1. The Control of 1 Train, 19, Section 101 1 to 1 May 6 at 21 May 10 May 6 at 21 May 10 May 10
```

Project Files:

```
■ deliveryTruckRoutes ~/Pych

✓ ■ CSV_files

       distances.csv
       locations.csv
       packages.csv
  > venv
     💪 data_loader.py
     hash_table_chaining.py
     locations.py
     the main.py
     nearest_neighbor.py
     packages.py
     truck.py
     table user_interface.py
> Illi External Libraries
  Scratches and Consoles
```

I1: Strengths of the Chosen Algorithm (Nearest Neighbor):

The nearest neighbor algorithm is a greedy algorithm in that it only looks at the following location's distance and cannot analyze future routes. In the worst-case runtime, the algorithm will perform poorly, but it typically performs well. The nearest neighbor algorithm is simple in that it will compare the current location to all other location distances. It will find the shortest distance

and travel to that location. It will repeat this process until all locations have been visited. In the program, I found that nearest_neighbor was not only simple but efficient in that the total distance traveled was much less than the limit of 140 miles; the total distance traveled was 105.2 miles.

I2: Verification of Algorithm:

The algorithm, nearest neighbor created routes for the trucks, which resulted in a total traveled 105.2 miles. All the packages were delivered on time and delivered according to their delivery specifications. The user interface provides all this data. When running the application and choosing option three, these results are shown. For packages that must be on truck two, they can be seen there, delivery times are shown, delayed packages are seen on truck two and leave at 9:05, and packages that must be delivered together can be seen on truck one. Truck three contains the package with the wrong address, and the results show that that package was delivered, and the path of truck three shows the new address.

13 & I3A: Other Possible Algorithms and Algorithm Differences:

Finding the most optimal solution for the traveling salesperson's problem is extremely difficult. One way to find the most optimal solution is to know every single possible route using a brute force algorithm. The use of a brute force algorithm where there are many possible routes becomes extremely slow. In use, it is not a good idea to use brute force because of the time complexity. Once the algorithm is solved, I'm sure it would be more efficient than the nearest neighbor, but the program would be much too slow. Another approach would be to use the nearest insertion algorithm, which has the same time complexity as the nearest neighbor. It would also be harder to implement as well. The algorithm starts at two locations and then repeatedly finds the closest location that has not been visited (Weru, 2019). It adds that location by placing the location between two of the closest locations that result in the shortest route (Weru, 2019). I found that it would be challenging to implement this algorithm in my code and that the time complexity does not make it any faster than nearest neighbor.

J: Different Approach:

If I were to rewrite the program, I would change the nearest neighbor algorithm to be more generalized. Looking back, I can see that my algorithm was made more specific for the current package data. I would have developed it now with a sense of future-proofing the program and have made it more general for different package data. Something that I found slightly challenging when cleaning up my code was that my original comments were not as helpful in explaining logic as they should be. Looking forward to future projects and applications, I will be more aware of the importance of proper commenting for all code that I write.

K1: Verification of Data Structure:

The graph data structure worked with the nearest neighbor algorithm to provide a total traveled 105.2 miles. All the packages were delivered on time and delivered according to their delivery specifications. The user interface provides all this data. When running the application and choosing option three, these results are shown. For packages that must be on truck two, they can be seen there, delivery times are shown, delayed packages are seen on truck two and leave at 9:05, and packages that must be delivered together can be seen on truck 1. Truck three contains the package with the wrong address, and the results show that that package was delivered, and the path shows the location of the new address. All package reporting (package status and

information) is available in the user interface by either searching for a specific package by ID or looking at option 3 in the interface. Packages can be efficiently retrieved by searching for the package id. When a package is searched for and found, all package components and their values are displayed.

K1A: Efficiency:

The hash table I chose to use in the program was a chain hash table. It is designed so that each index in the hash table is a list and that the object being searched for will return when finding the correct item while looping through the list. However, the program was designed so that only one element is stored in each index of the hash table. Storing only one element in each hash table index makes it more efficient, O(1), when searching, inserting, and removing data. If more packages were to be added, not through the user interface, then chaining would take place, and searching, inserting, and removing data would be the worst-case runtime of O(N).

K1B: Overhead:

Adding more packages to the hash data structure would require more space in the computer's memory. Larger package input sizes result in more usage of memory.

K1C: Implications:

Adding more cities would not affect the look-up time and space usage of the hash table in the program. Trucks are stored in a separate hash table than the package hash table, and the size of the hash table is four for indices 0, 1, 2, and 3. Truck 1 is in index 1, truck 2 in index 2, and truck 3 in index 3. Adding more trucks would use the chaining feature in the algorithm, and the worst-case runtime would be O(N), whereas it is currently O(1); to keep the hash table constant, creating a larger hash table that has one truck at each index would be more efficient. Locations are stored in graphs and would not affect the hash table performance.

K2: Other Data Structures:

Two other data structures that could store package data instead of the chain hash table are the linear probing hash table and a list. The linear probing hash table implementation would be efficient just like the chain hash table and would be a great option. The list would work as well, each package could be a list within the list, and the packages can be searched for by ID with a function if that option was used for storing packages.

K2A: Data Structures Differences:

When comparing the chain hash table and the linear probing hash table, the linear probing hash table is more complex in searching, updating, and removing items. There are empty buckets since the start and empty after removal buckets in the hash table. The linear probing hash table would work just fine, but the simplicity of the chain hash table was more appealing to me. Also, the time complexities of the different hash tables are identical, and neither option was better than the other. The list, on the other hand, would be much slower than an optimized chain hash table. The chain hash table in the program performs operations at best case in O(1) time and worst (unlikely) in O(N) time. The list also performs best case in O(1) time and worst-case O(N) time. However, the chain hash table in the program has been designed to have only one object in each buck, so it performs at O(1) time all the time. The list would be slower and also be more complex in completing searching, updating, and deleting.

L: Sources:

Weru, L. (2019, December 28). 11 Animated Algorithms for the Traveling Salesman Problem.

STEM Lounge; STEM Lounge. https://stemlounge.com/animated-algorithms-for-the-traveling-salesman-problem/