

# Route Optimization using Reinforcement Learning and Metaheuristic Approach

publicis  
sapient

September 8, 2022

## About the Speakers



Ravi Ranjan is working as Manager Data Science at Publicis Sapient (India) with expertise in building scalable ML solutions. He is a certified Google Cloud Architect. He has done proficiency course in Reinforcement Learning from IISc, Bangalore. He is contributor and member at Kubeflow (ML Platform by Google).

publicis  
sapient



Sushant has a bachelor's in Materials Engineering from the Indian Institute of Technology, Kharagpur. He has extensive work experience in Reinforcement Learning, Computer Vision with AR technologies and created end-to-end pipelines and data products from conceptualization to deployment phase for various engagements. He is currently working as a Senior Data Scientist at Publicis Sapient.

## Session Logistics

1. Access to the session environment using the following link. [<https://bit.ly/join-the-ODSC-session>]
2. Presentation and code will be available at link. [<https://bit.ly/ODSC-conference-India-2022>]
3. Connecting to the speaker [Please send introductory note in LinkedIn invite]  
 <https://bit.ly/ravi-ranjan-03>  
 <https://bit.ly/sushant-burnawal>
4. Don't forget to tweet and share the session with **#ODSCAPAC**

## Learning Outcome

1. Gain an understanding of Multi-Objective Optimisation .
2. How to transform the concepts and build Route Optimizing Module?
3. Gain a detailed understanding of deep reinforcement learning and Metaheuristics based Optimisation
4. Check out the use of open-source software and other scaling solutions for the Route Optimizer Module
5. Reference architecture of the Route Optimizer using and Veroviz

# Session Agenda

1. Introduction to Vehicle Route Problem
2. Why we need Route Optimization?
3. The classical approach to Multi-vehicle route optimization and its limitations
4. Introduction to the key concepts of
  - Multi-objective Optimization
  - Reinforcement Learning
  - Genetic Algorithms and other Metaheuristics
4. Combination of a Deep RL-based algorithm with a metaheuristic approach
5. Training Methodology and Result Discussion
6. Business Impacts and Outcomes
7. Question - Answers

Section 1

# Vehicle Routing in Logistic Industry

# Logistic Industry - World Insight

8.6tr  
USD

12.9tr  
USD

6.5%

52%

Market Size in 2022

Projected Market Size in  
2027

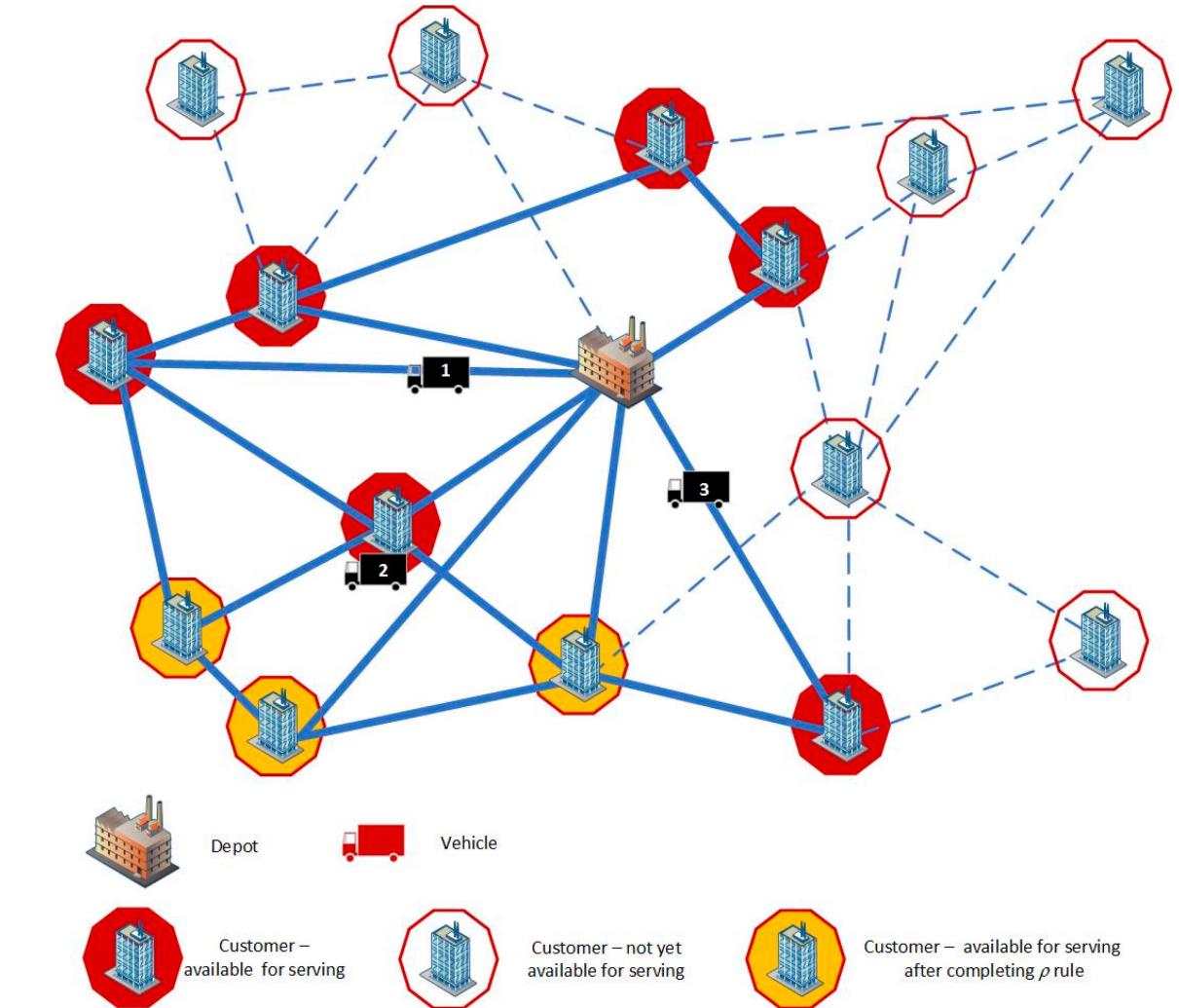
Of CAGR

Of Global  
Transport Model  
is Roadways

The background to our problem is...



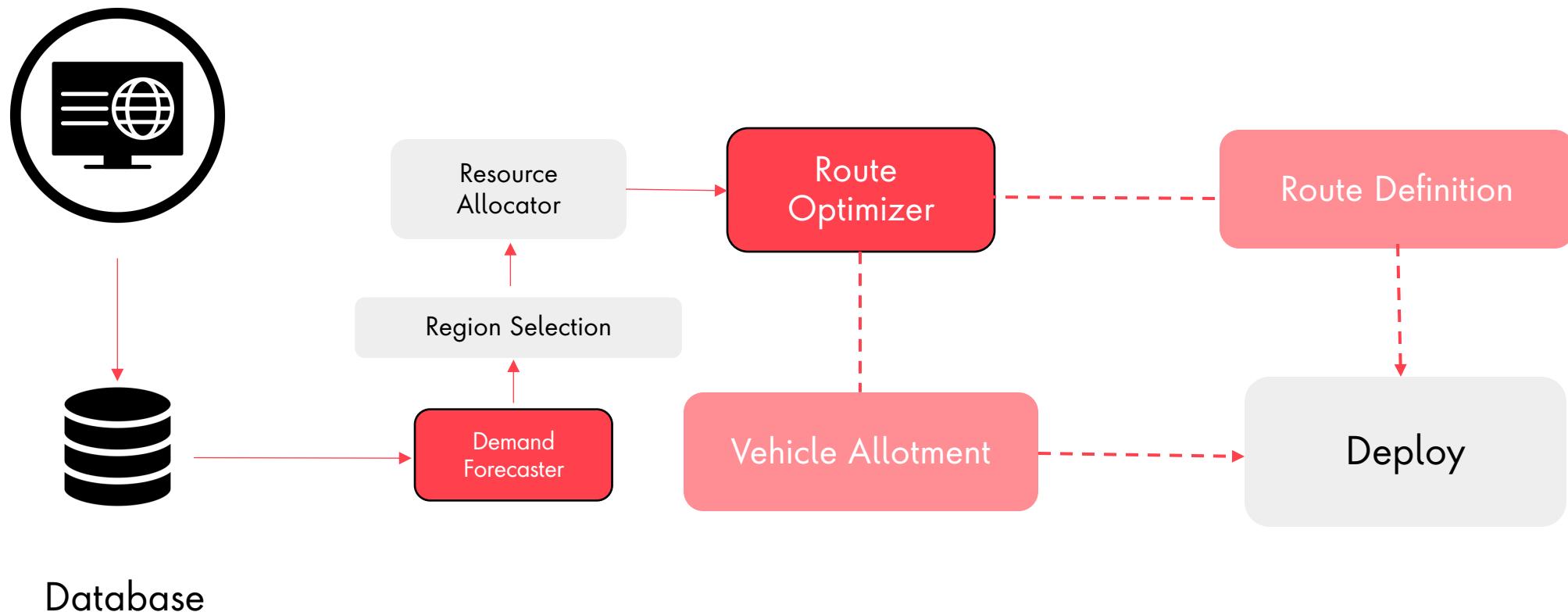
# Vehicle Routing Problem



Section 2

# Route Optimization

## The business flow...

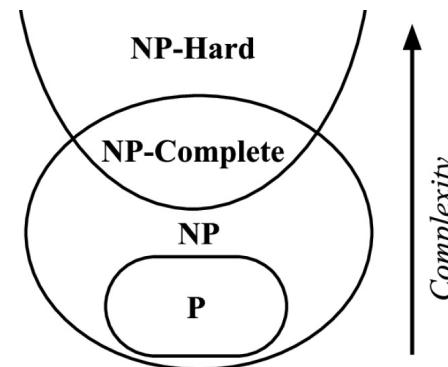


*Daily based Model*

## Classifying the decision problems into the complexity classes

P(Polynomial Time)

NP(Non-deterministic  
Polynomial Time)



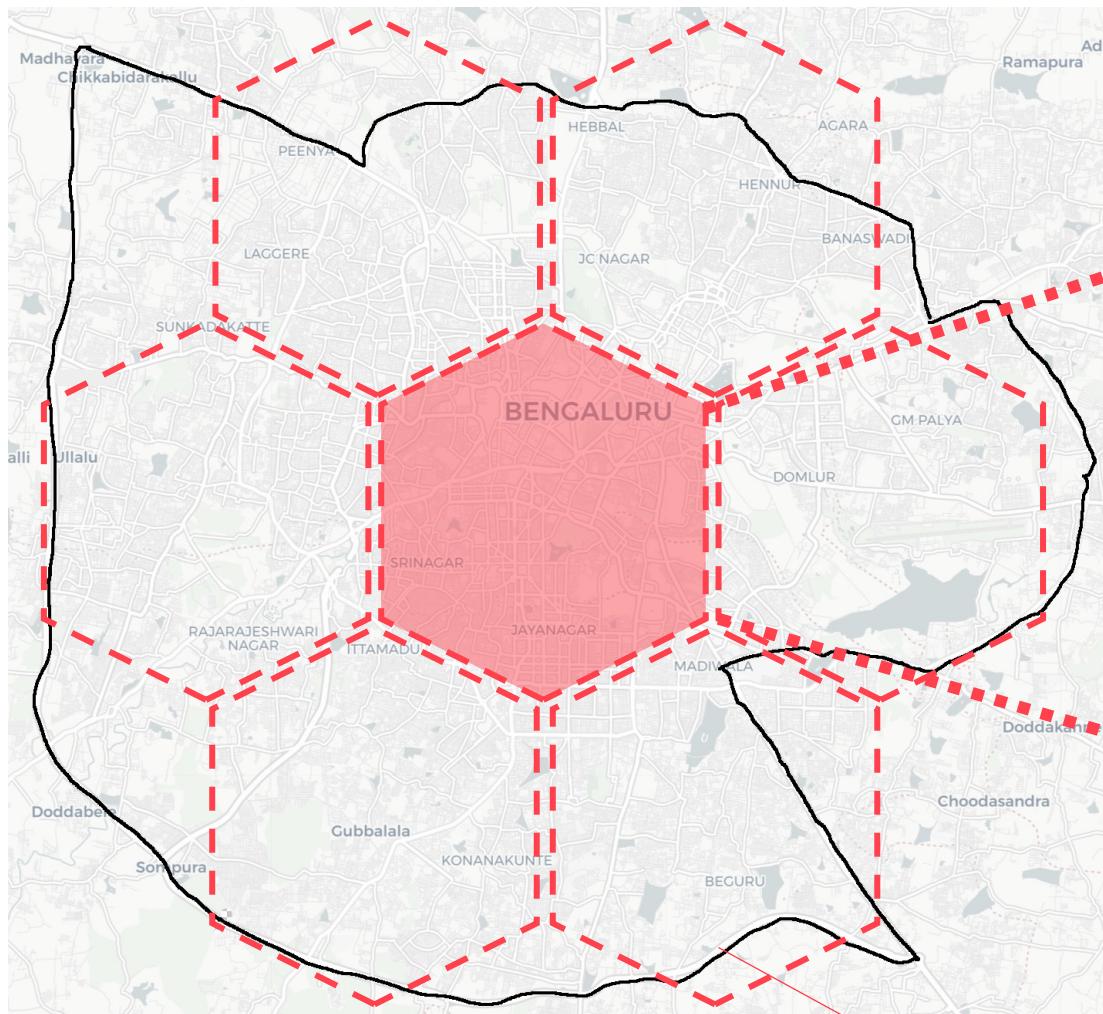
NP-Hard

NP-Complete

Section 3

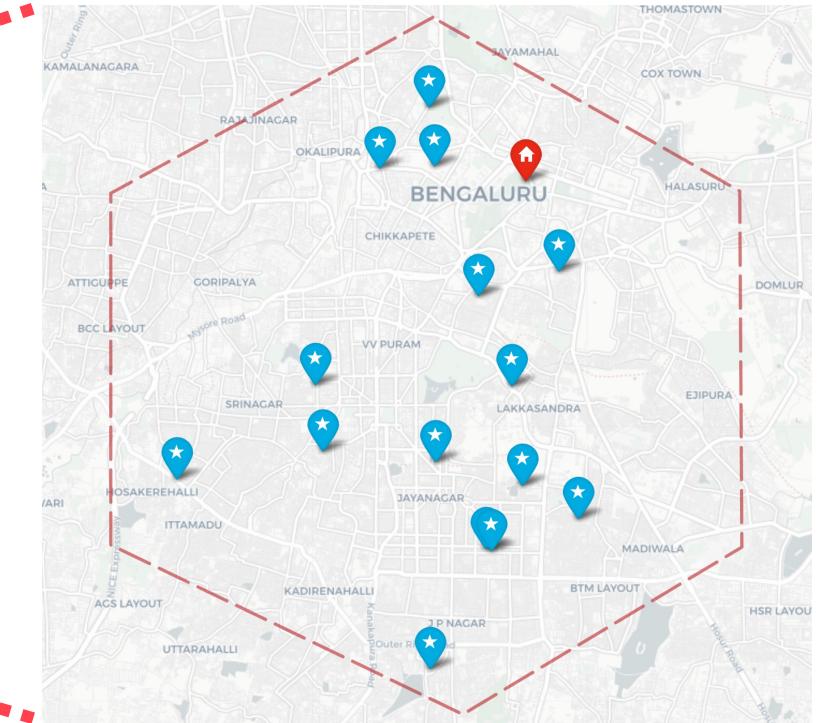
# Scaling the solution using Primary Cells

## Creating a Scalable Architecture...



Outer ring road

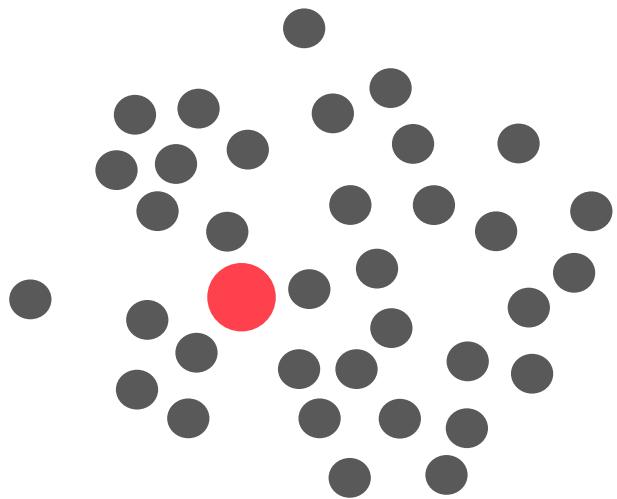
## Central Bangalore



Section 4

## Route Optimizer Module - Explanation

## Problem we are facing now...classic derivation of NP-hard



- Depot →  $n = 0$
- Customer →  $n, weight = 1, w_1$   
                   $2, w_2$   
                  .  
                  .  
                  .  
                   $k, w_k$

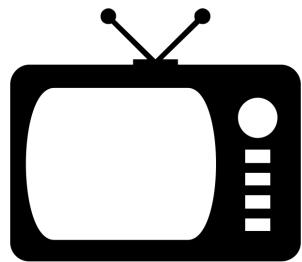
Set of customers along with their demands



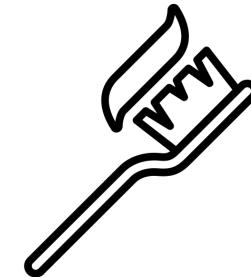
*(Q) : Each vehicle has a fixed Capacity*

"Television" v/s "toothbrush" v/s "cabbage"....

- ◇ Each object is allotted a corresponding size e.g.



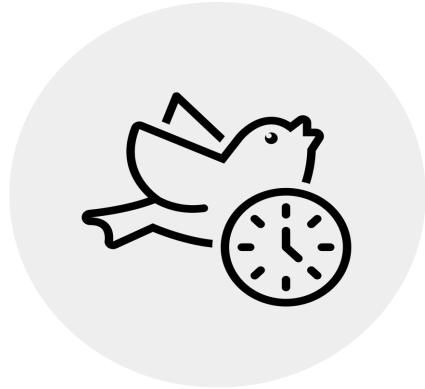
size = 90 units



size = 1 unit

## Time Durations to keep in mind ...

### Time Window-early



The earliest time that a vehicle can arrive

### Time Window-late



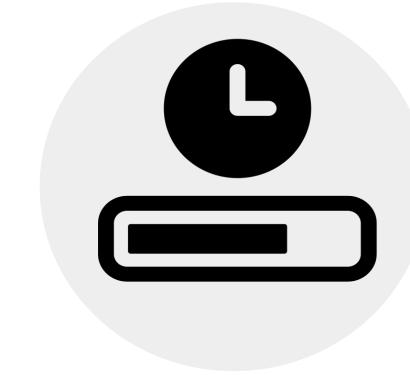
The latest time that a vehicle can arrive

### Time Window-service Time



Time needed to unload the goods

### Time Window - wait cost



Cost for waiting for a client to be available

## Types of vehicles in use...

Store-owned vehicles



Rented vehicles



Size

To satisfy average demand

To satisfy large demand

Fixed Cost

Lower

Higher

Variable Cost

Lower

Higher

Velocity

Same

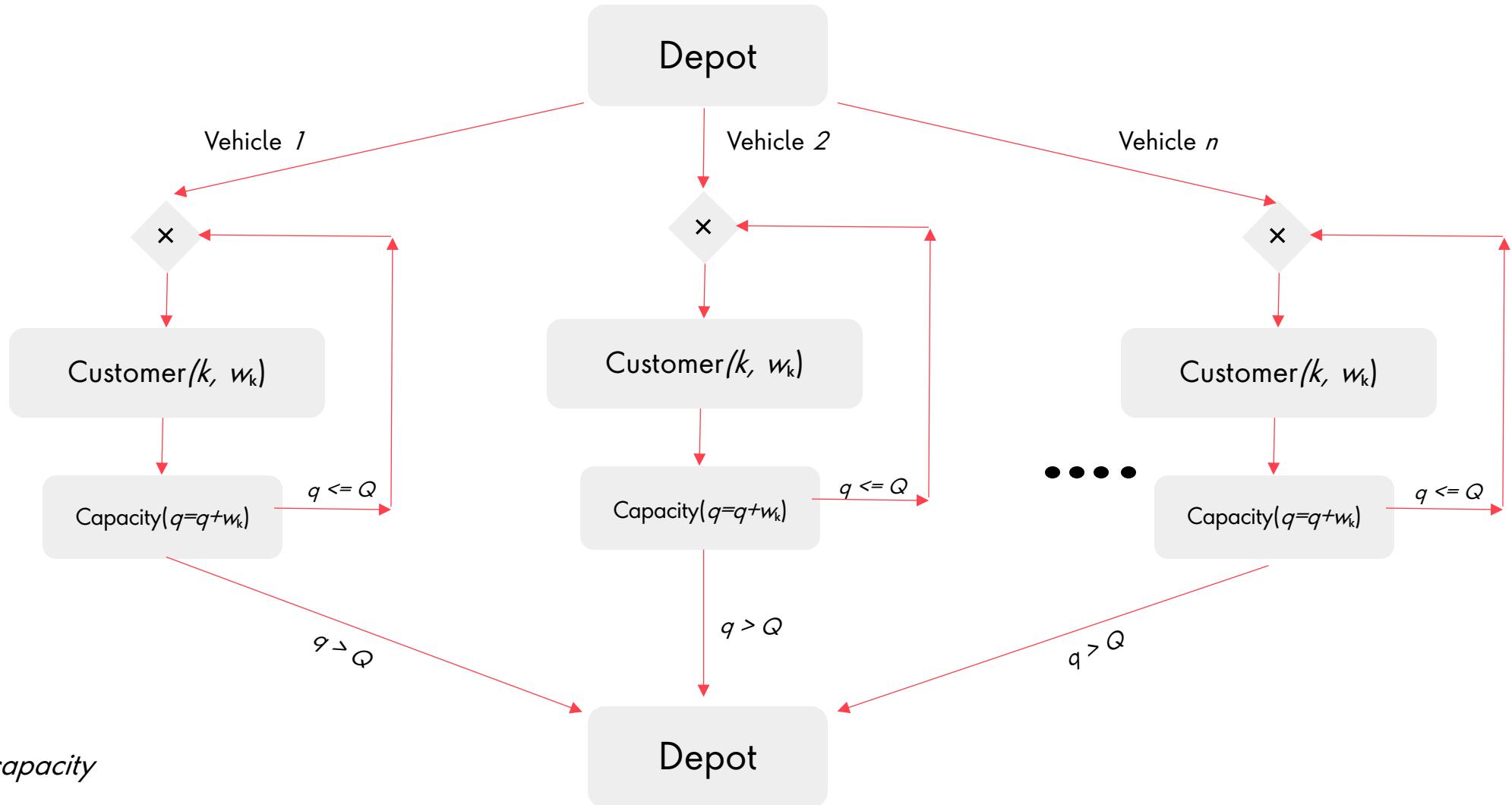
Same

Fleet Size

Same

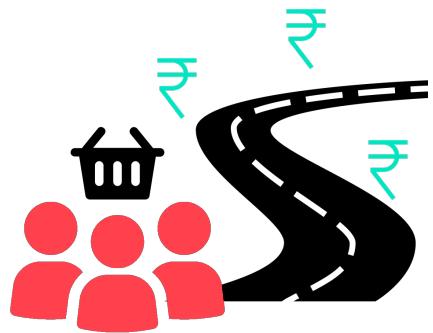
Same

Continuing with the detailed workflow...

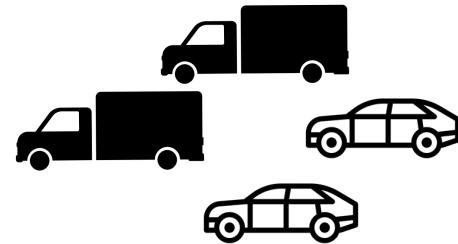


## The objectives...

- ◆ Optimization has been performed on the following two objectives:

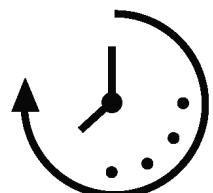


*The cost to deliver*

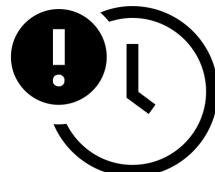


*Number of vehicles*

- ◆ Objectives omitted :



*Total time*



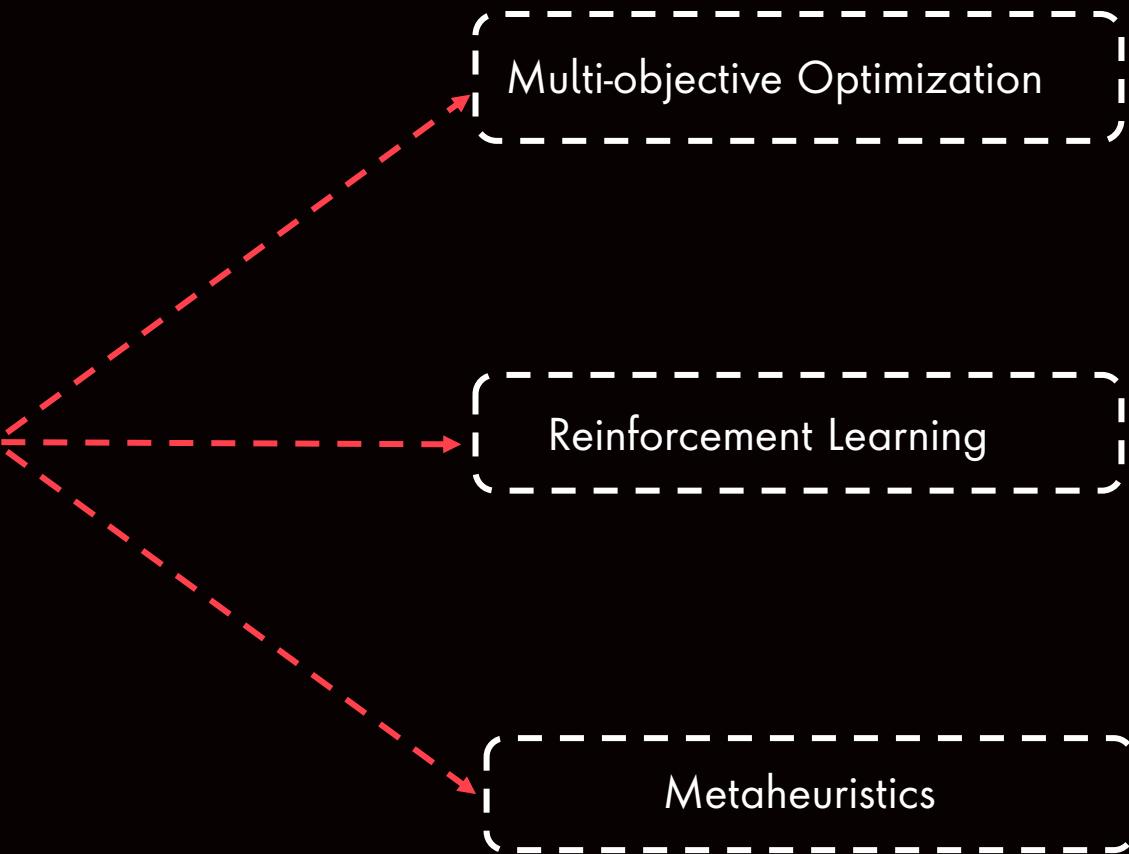
*Time delay*

Section 5

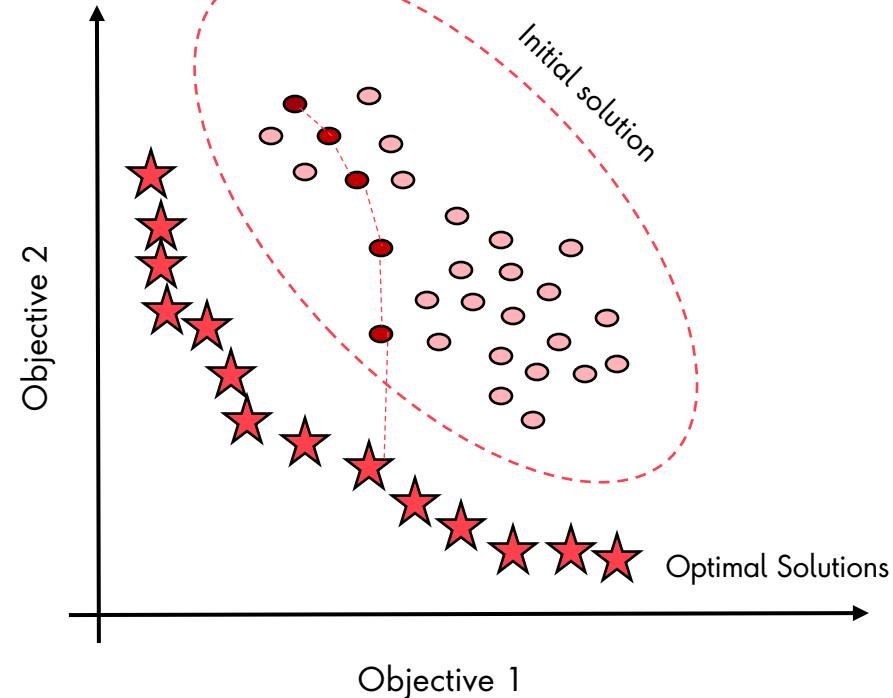
## Route Optimizer Module - Demo

Section 6

# Key Concepts

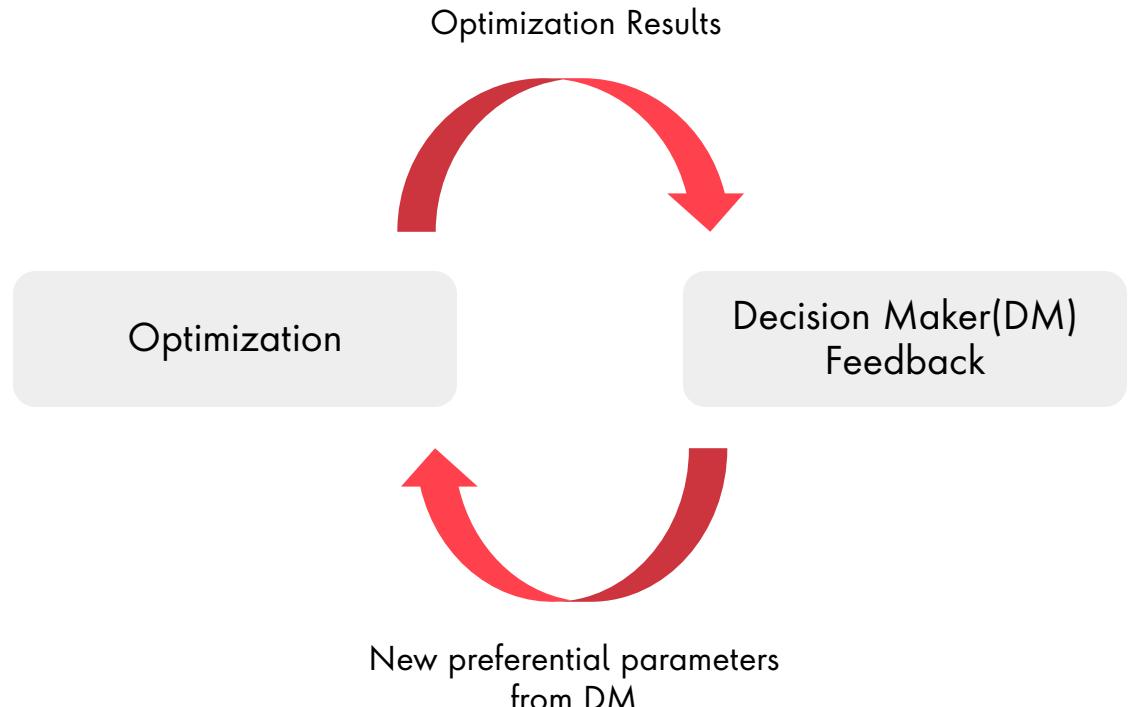


## Introduction to multi-objective optimization

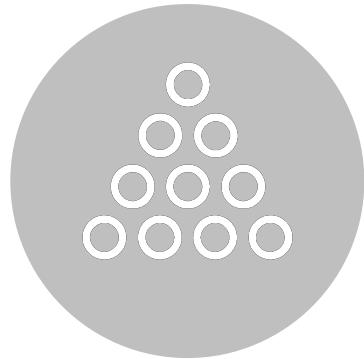


○ - Dominated Solution

★ - Non-Dominated Solution



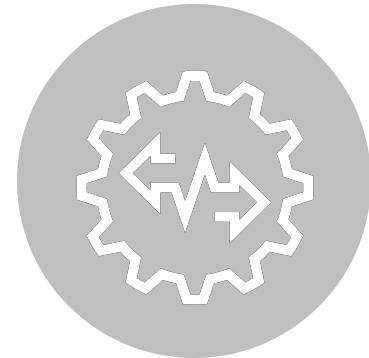
## Why we require multi-objective optimization ?



More than one optimal solution



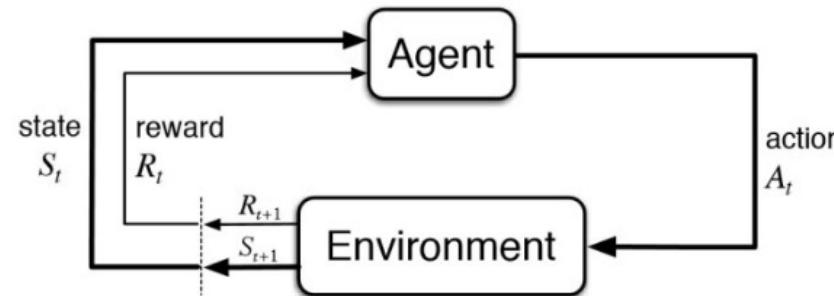
Real-World Engineering Problems



Conflicting Objectives

## Introduction to Reinforcement Learning

Reinforcement Learning(RL) is a goal based learning, based on interaction from the environment. It is an approach to teach machines to interact with environments and receive rewards for performing the right actions until their goal is achieved.

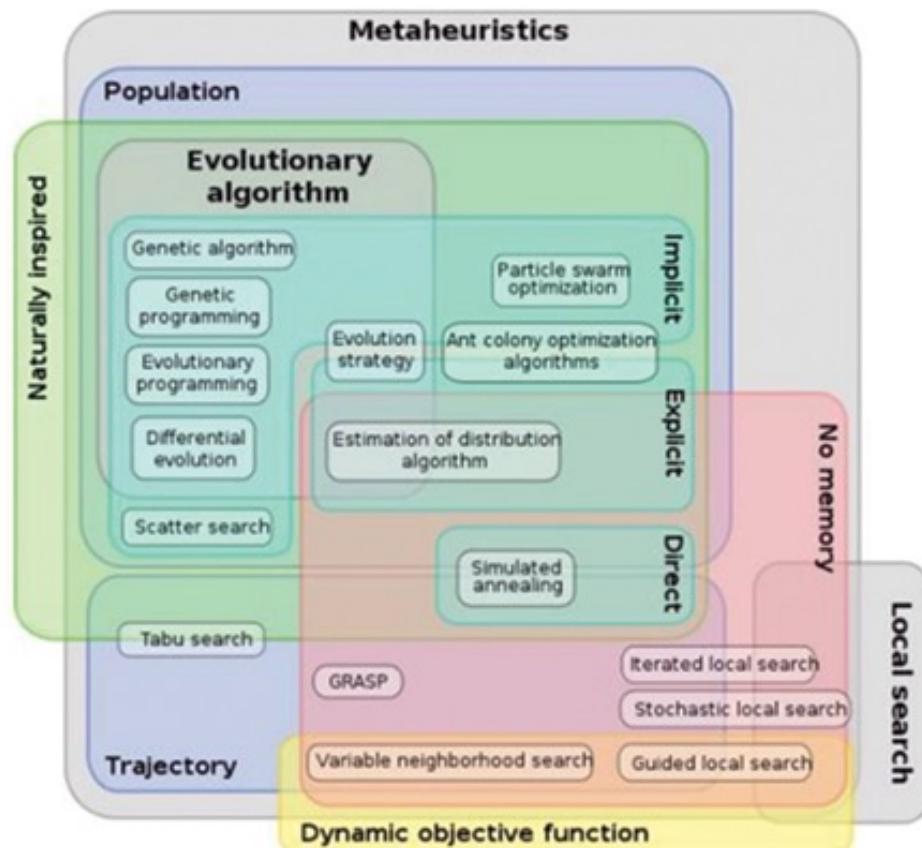


Some key terms describing the elements of a Reinforcement Learning Problem are:

- **Environment** : World in which the agent operates containing the depot, nodes, vehicles and rules governing there operation
- **State**: An immediate situation in which the agent finds itself in
- **Action**: It's the steps of choosing the next node recommended to a user to maximise rewards
- **Reward/Penalty**: The feedback which the agent gets hereby determined by negative of the total distance travelled
- **Policy**: Strategy employed by the agent to minimize the loss objective while satisfying the constraints of time and capacity

## Introduction to metaheuristics

Derived from the 'meta' and 'heuristic', it's a high-level methodology used for finding new strategies for solving a problem. These approaches can be distinguished into following types :



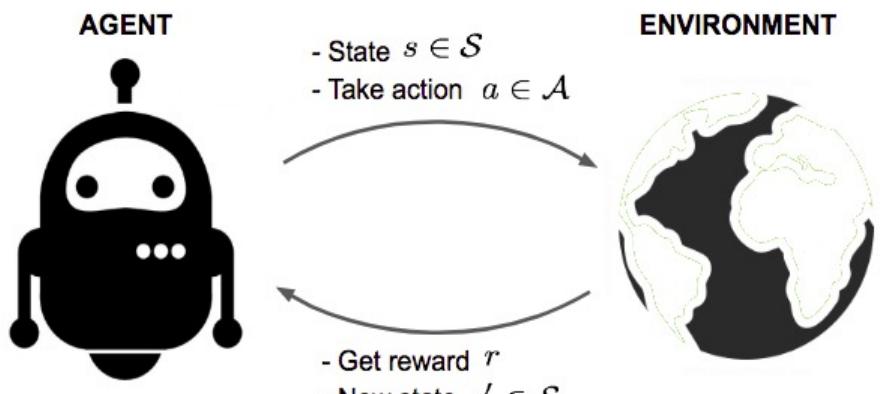
Here we will be looking into the Population based approaches namely :

- Genetic Algorithms
- Ant Colony Optimization

Section 7

# Combination of RL and Metaheuristics

## Details of the Model Building



$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \text{ where } \gamma \in [0, 1)$$

$$R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} \dots$$

---

**Algorithm 4** Asynchronous Advantage Actor-Critic (A3C)

---

```

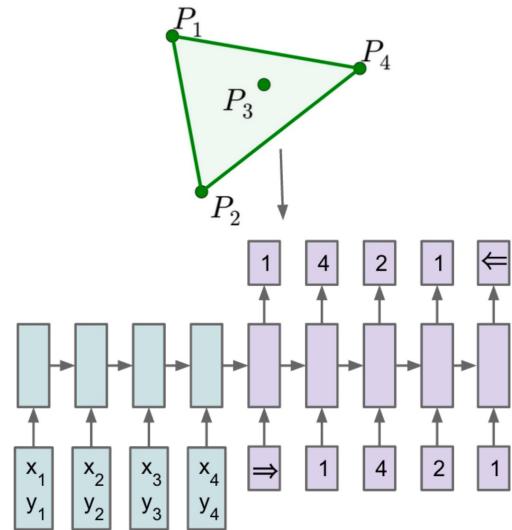
1: initialize the actor network with random weights  $\theta^0$  and
   critic network with random weights  $\phi^0$  in the master
   thread.
2: initialize  $N$  thread-specific actor and critic networks
   with weights  $\theta^n$  and  $\phi^n$  associated with thread  $n$ .
3: repeat
4:   for each thread  $n$  do
5:     sample a instance problem from  $\Phi_M$  with initial
       state  $X_0^n$ 
6:     initialize step counter  $t^n \leftarrow 0$ 
7:     while episode not finished do
8:       choose  $y_{t+1}^n$  according to  $P(y_{t+1}^n | Y_t^n, X_t^n; \theta^n)$ 
9:       observe new state  $X_{t+1}^n$ ;
10:      observe one-step reward  $R_t^n = R(Y_t^n, X_t^n)$ 
11:      let  $A_t^n = (R_t^n + V(X_{t+1}^n; \phi) - V(X_t^n; \phi))$ 
12:       $d\theta^0 \leftarrow d\theta^0 + \nabla_{\theta} A_t^n \log P(y_{t+1}^n | Y_t^n, X_t^n; \theta^n)$ 
13:       $d\phi^0 \leftarrow d\phi^0 + \nabla_{\phi} (A_t^n)^2$ 
14:       $t^n \leftarrow t^n + 1$ 
15:    end while
16:  end for
17:  periodically update  $\theta^0$  using  $d\theta^0$  and  $\phi^0$  using  $d\phi^0$ 
18:   $\theta^n \leftarrow \theta^0, \phi^n \leftarrow \phi^0$ 
19:  reset gradients:  $d\theta^0 \leftarrow 0, d\phi^0 \leftarrow 0$ 
20: until training is finished

```

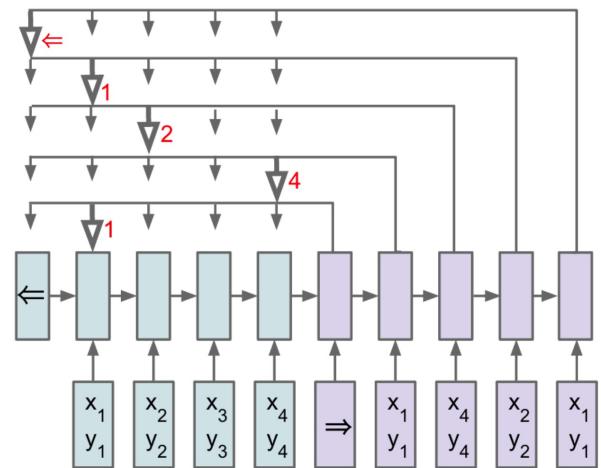
---

# Why Pointer Networks instead of Sequence-to-Sequence ? a modified form of the Pointer networks

- 1.The output of pointer networks is discrete and correspond to positions in the input sequence
- 2.the number of target classes in each step of the output depends on the length of the input, which is variable.



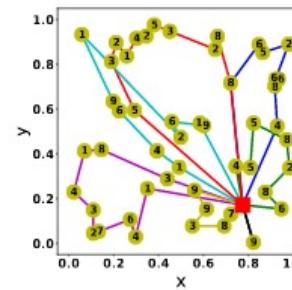
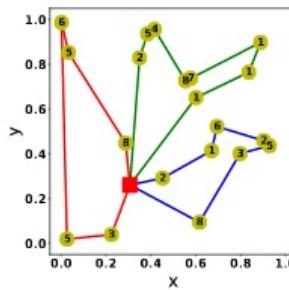
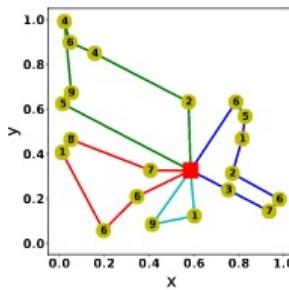
(a) Sequence-to-Sequence



(b) Ptr-Net

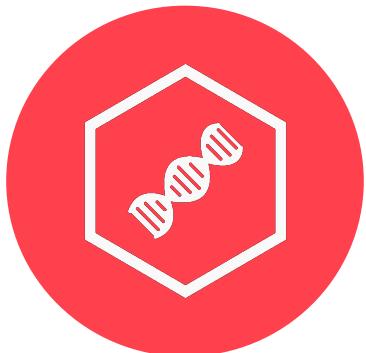
## Training methodology and Results

### Results



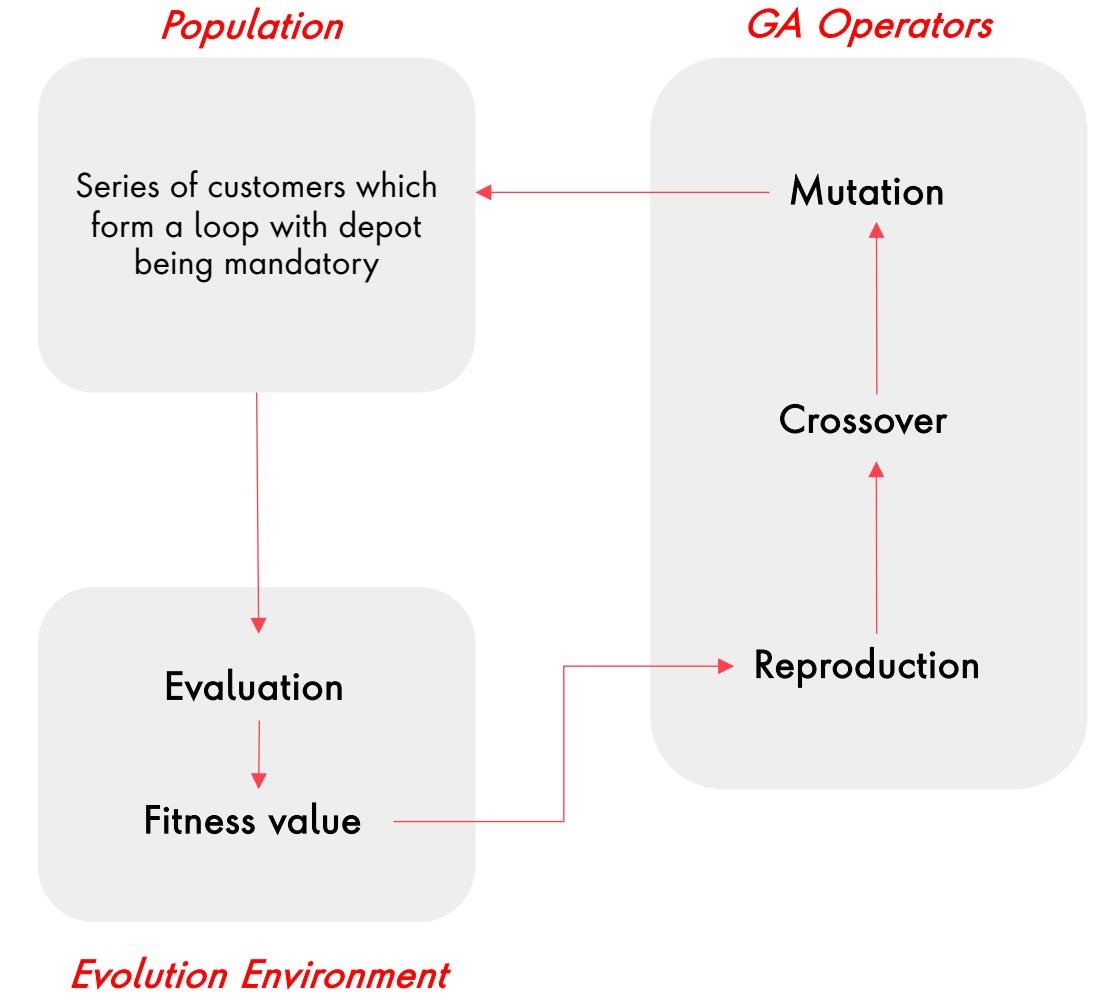
Here value written on each node  
is weight of the node

## Combining with Metaheuristic Approaches - I

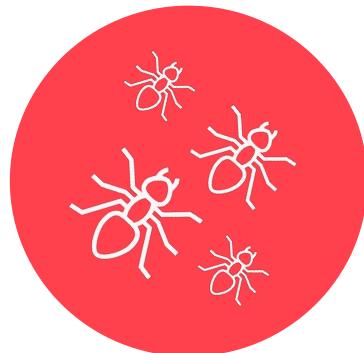


**Genetic Algorithm**

Based on Survival of the fittest  
Steps involved are -  
**Initial Population, selection, crossover, mutation**



## Combining with Metaheuristic Approaches - II



### Ant Colony Optimization

Ant go through food while laying down pheromone trails which attract other ants to follow same path.  
Shortest path, more pheromone trail

### The determining quantities of ACO:

- Ant Density
  - Ant Quantity
  - Any Cycle
- Pheromones updated in each movement
- Pheromones updated after all ants completed their tour

### Conditions of Pheromone Update:

$$\tau_{xy} \leftarrow (1 - \rho)\tau_{xy} + \sum_k \Delta\tau_{xy}^k$$

$$\Delta\tau_{xy}^k = \begin{cases} Q/L_k & \text{if ant } k \text{ uses curve } xy \text{ in its tour} \\ 0 & \text{otherwise} \end{cases}$$

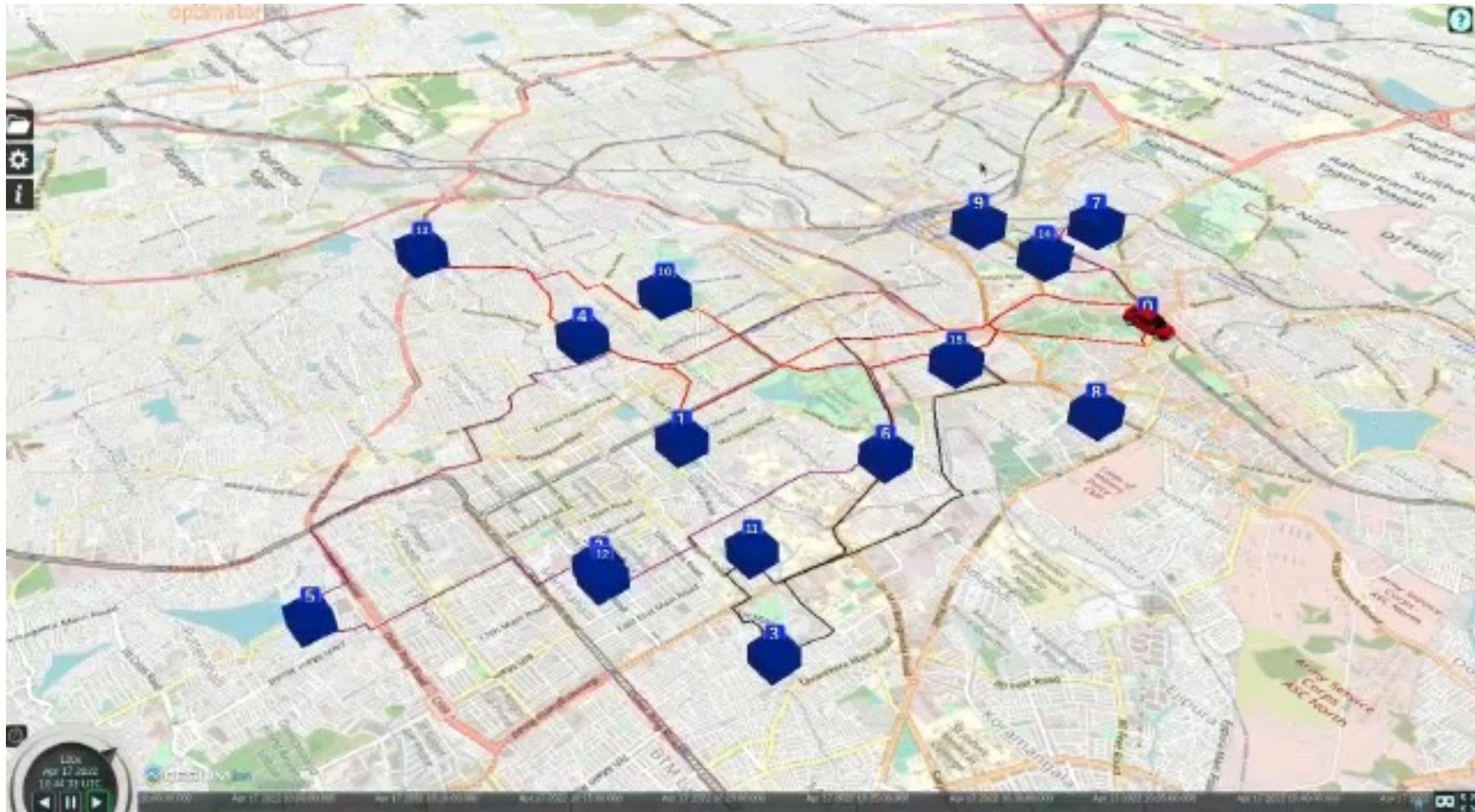
Section 8

# Training and Result Discussion

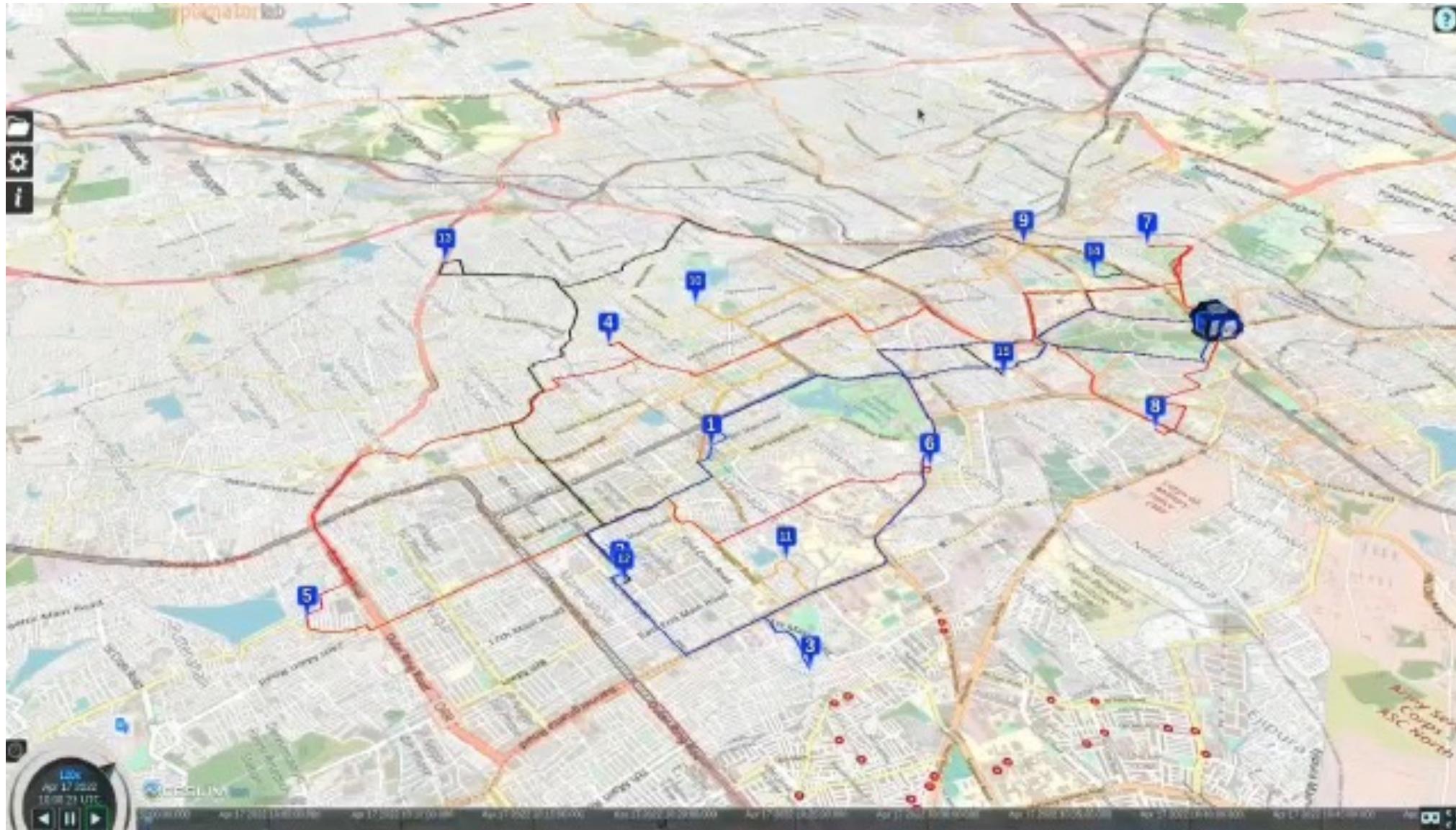
## Comparison between the Base and Proposed Model...

	Genetic Algorithm	Baseline
Cost	₹ 1557	₹ 2882
Number of Trucks	3	4
Number of Cars	1	0
Distance	87 km	116km

## Visualization of Route Optimization in Central Bangalore - Proposed



## Visualization of Route Optimization in Central Bangalore - Baseline



Section 9

# Future Works - Handling the Anomalies

## How to deal with Weather Anomalies – using the weather API



Bad weather at a location?



Increased wait cost

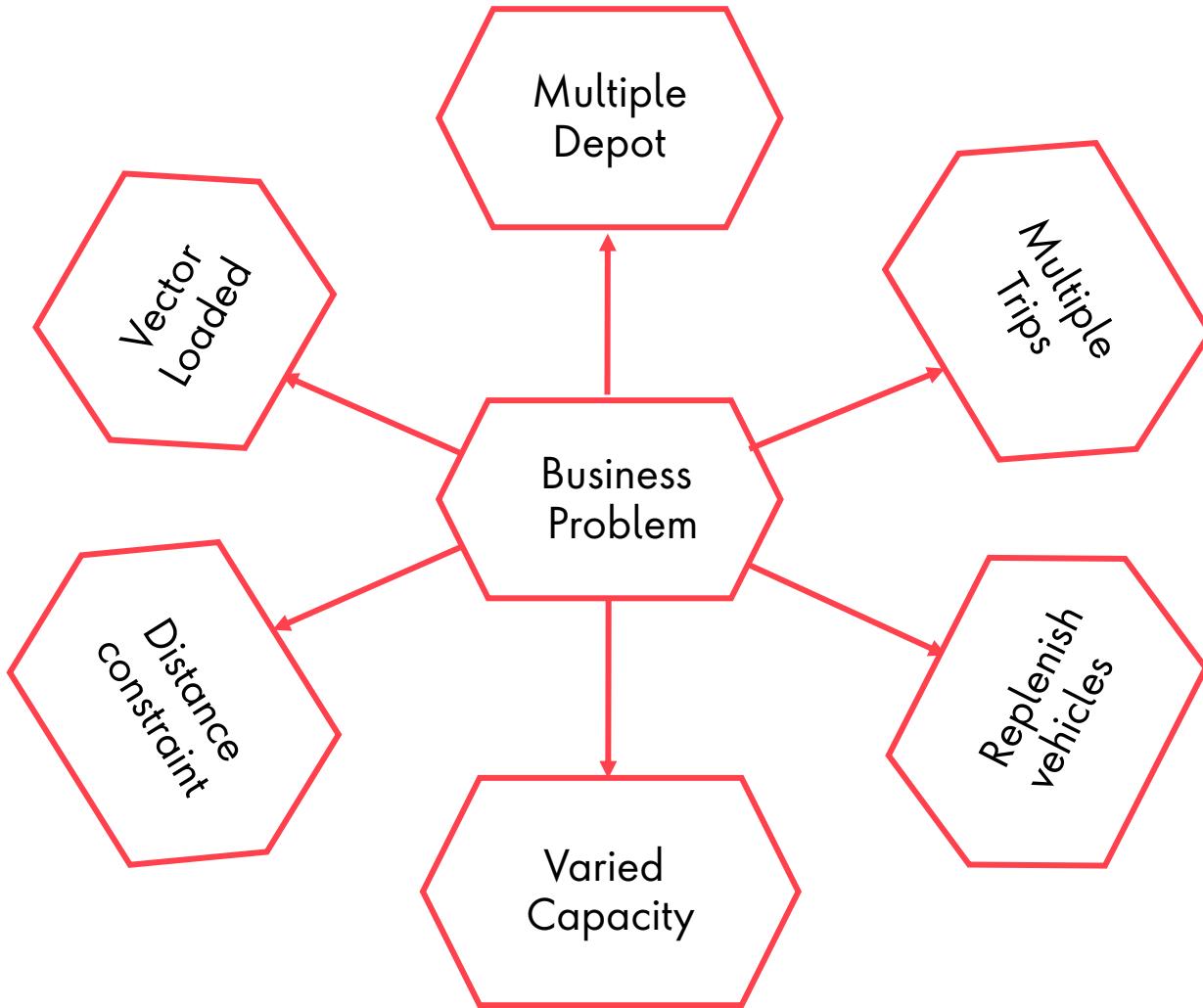


Increased service time

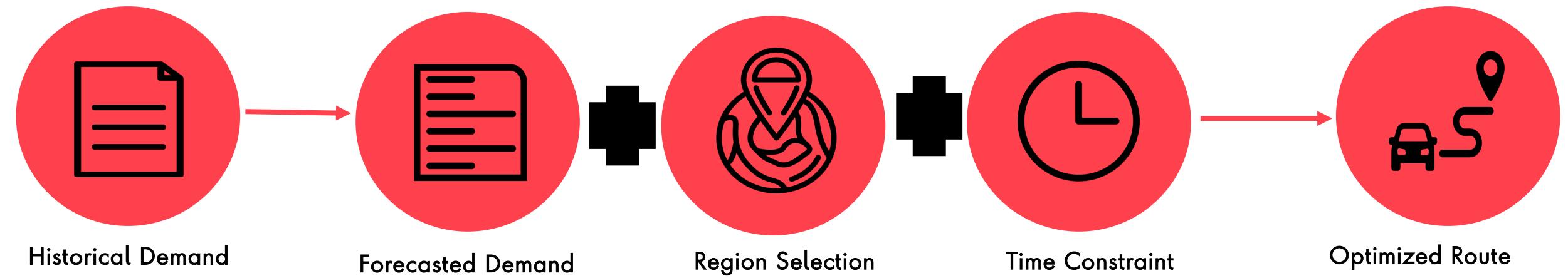


Weather API delivers the information based on longitude , latitude of a location

## Extensions to our business problem



## Summary



thank you