Modeling the Stream Rate of Vehicular Networks for Predictable Auto-Scaling of Edge-Cloud Systems

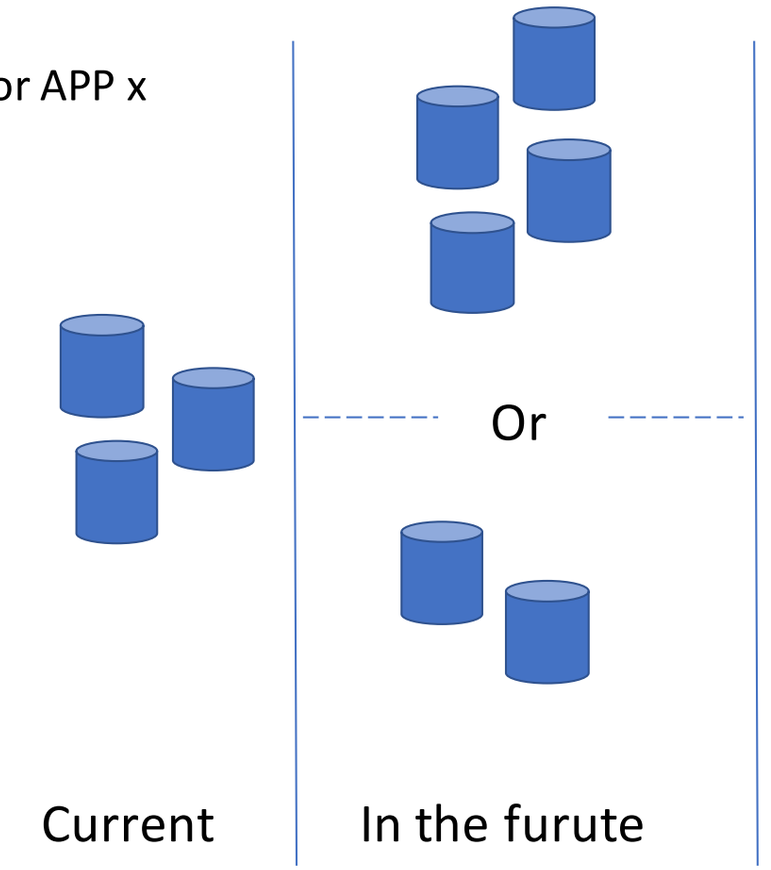
Edgar ROMO, NDS Laboratory, CIC-IPN
KerData Team, INRIA

Introduction

- **Context:** Vehicular Networks connected to the Cloud

- **Objective:** take advantage of the Edge and Cloud processing capabilities efficiently

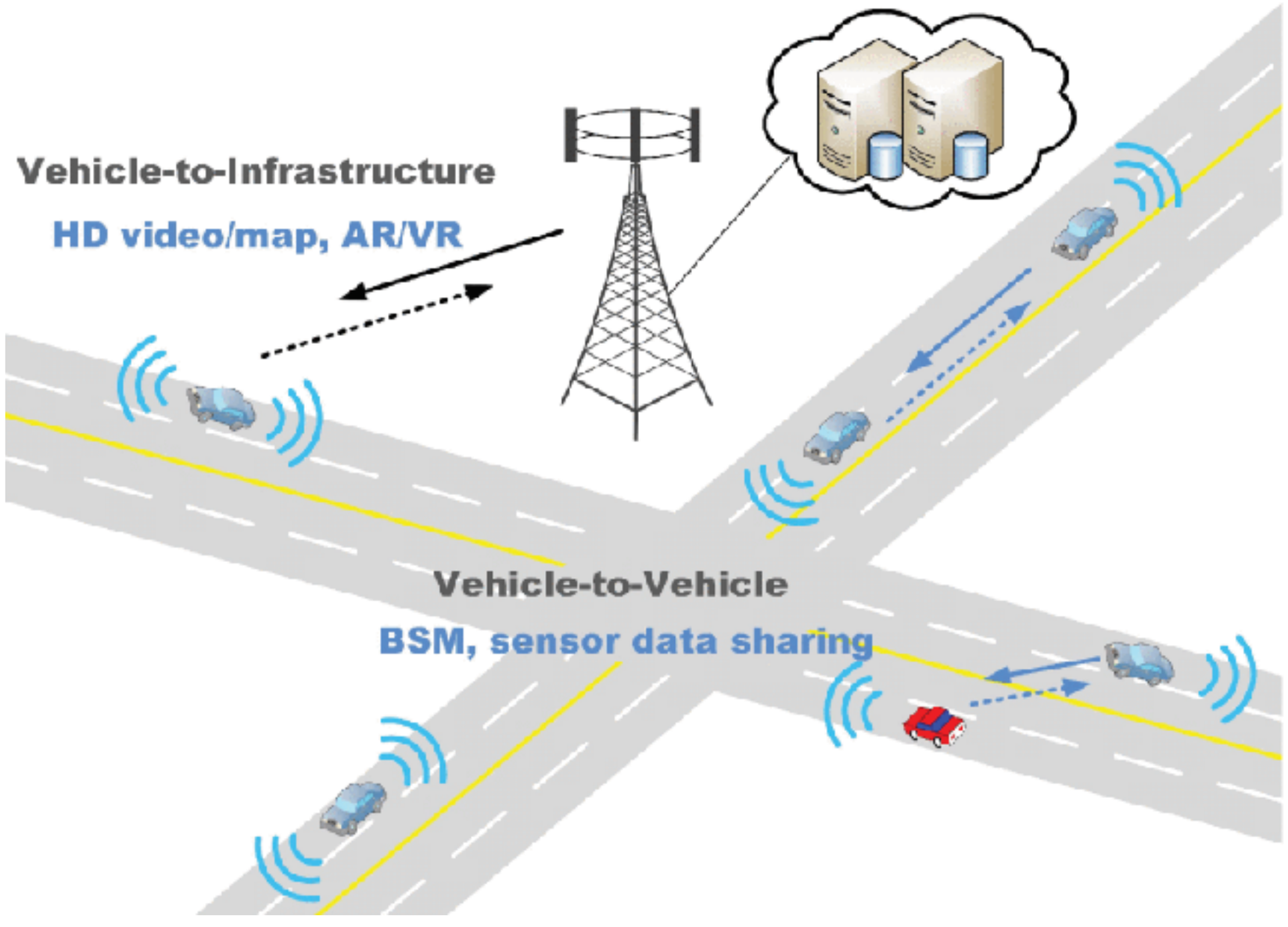
- **Use-case:** auto-scaling for Edge-Cloud architectures



Goals

- Modeling the stream arrivals rate
  - Use of Edge Services to predict the load of information to be processed
  - Making a more realistic prediction using Coxian distributions
  - Mathematical model for arrival of information
- Designing an Auto-Scaling module to reserve resources in the Cloud
  - Reduction the amount of information processed in the Cloud

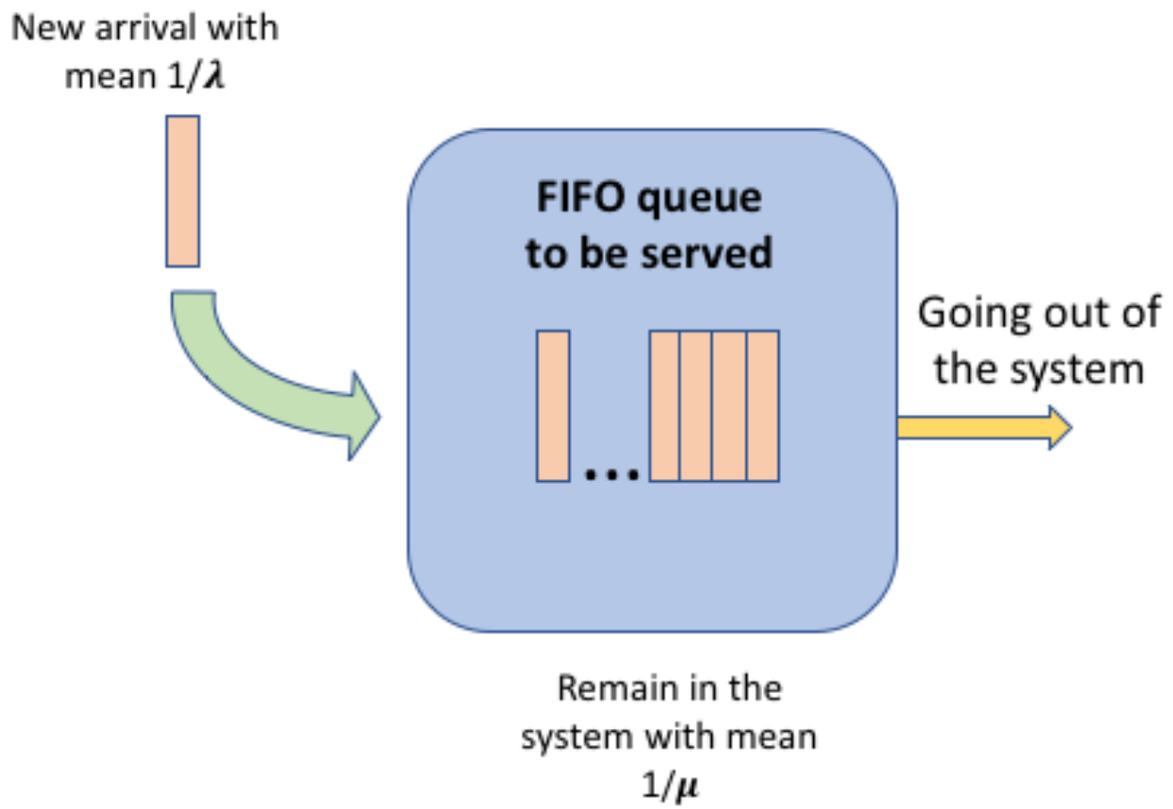
Vehicular Networks



* Taken from Le Liang, Toward Intelligent Vehicular Networks: A Machine Learning Framework

Prediction of the arrivals

- Classical arriving/service time model for networks based on queue theory
- Usage of exponential distribution to model times of service



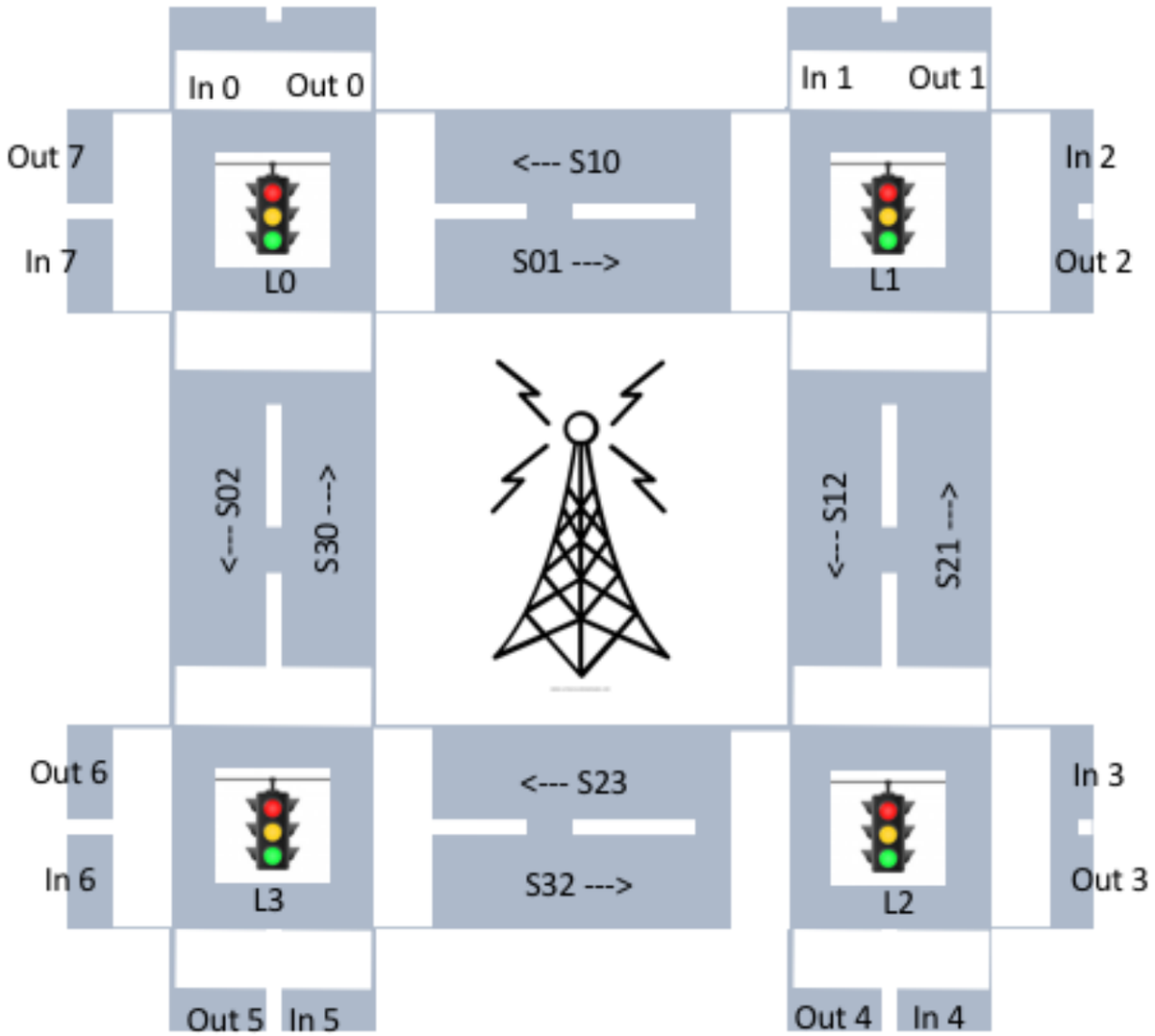
Prediction of the arrivals

- Another approach is to see like a black box where vehicles enters and travel inside a coverage area

- Inside the coverage area, different situations could occur

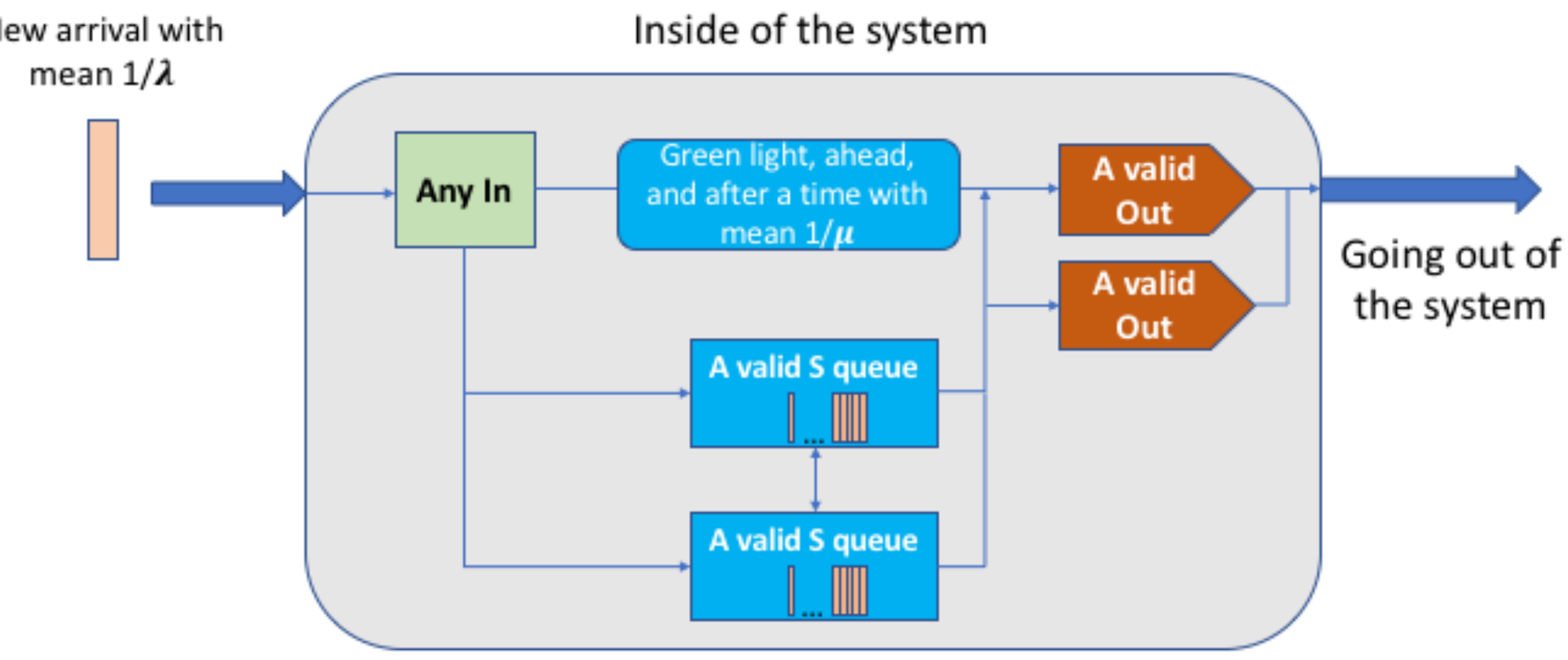
Prediction of the arrivals

- Predicting model based on traffic lights
- Coverage area where vehicles move inside it by a random time



Prediction of the arrival

- Although multiple random variables with exponentials distribution are considered, the time inside the coverage area is not a exponential distribution

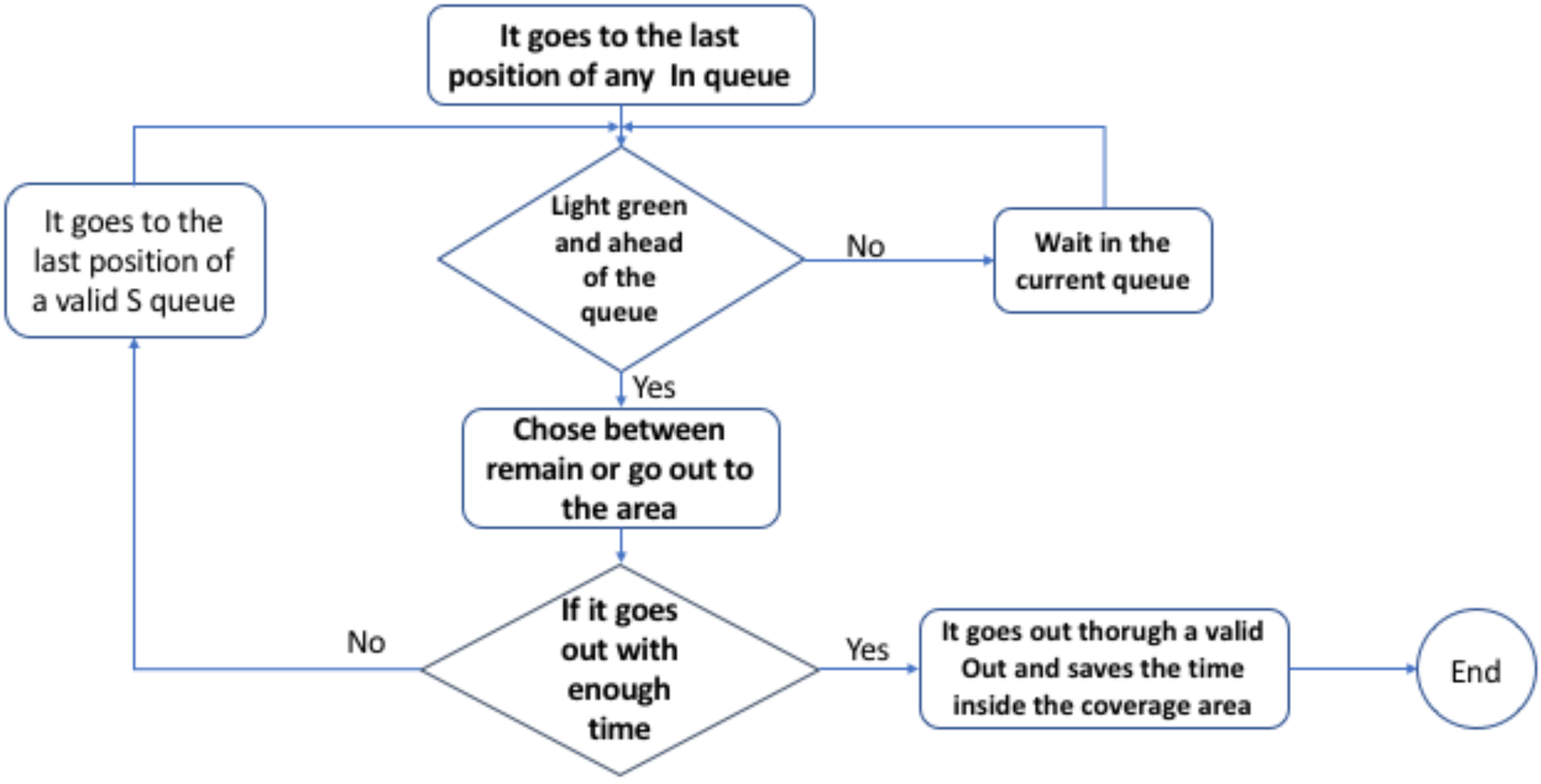


In any S queue, movements only with green light, first in the queue and mean time $1/\mu$

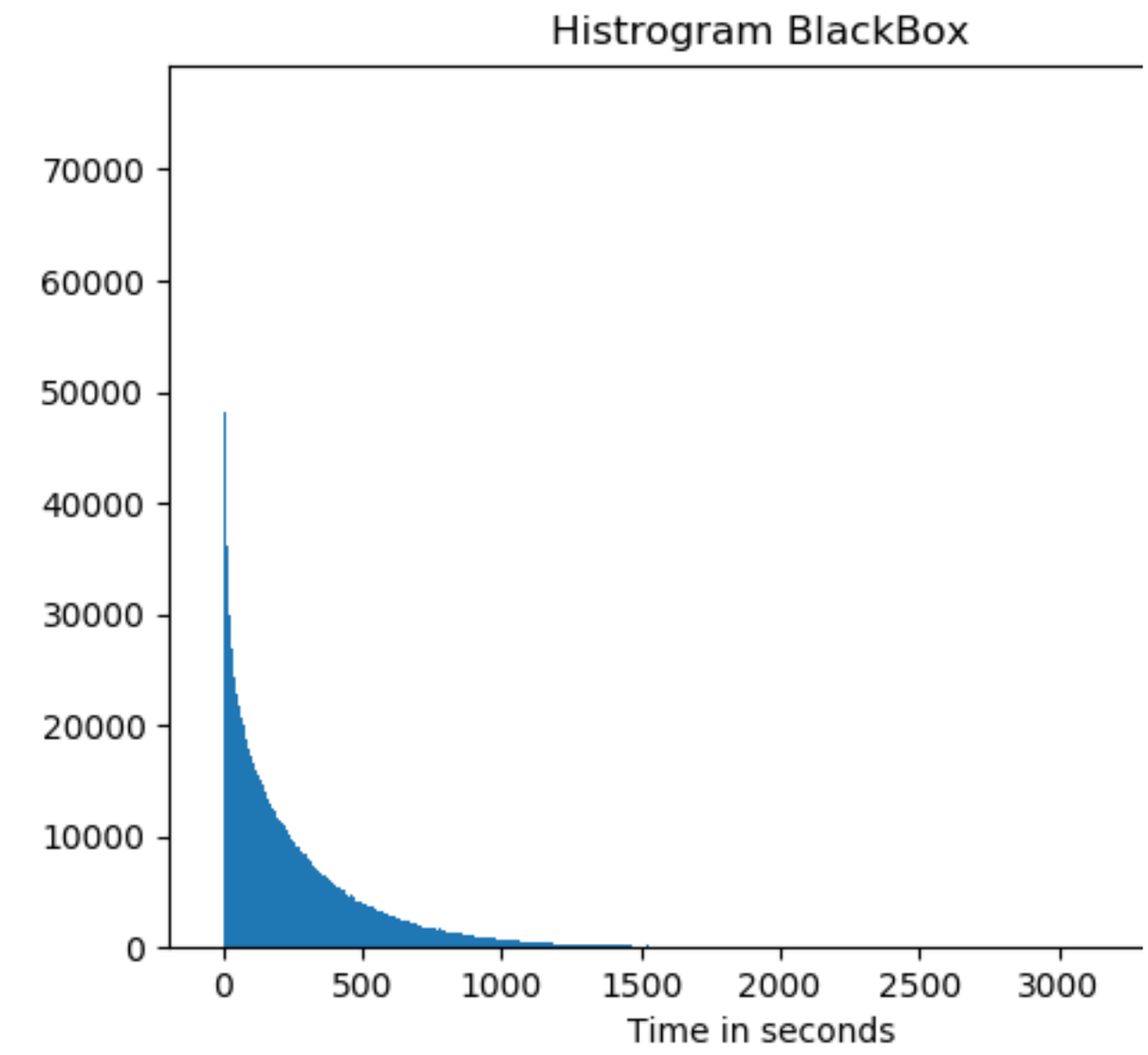
Assumptions

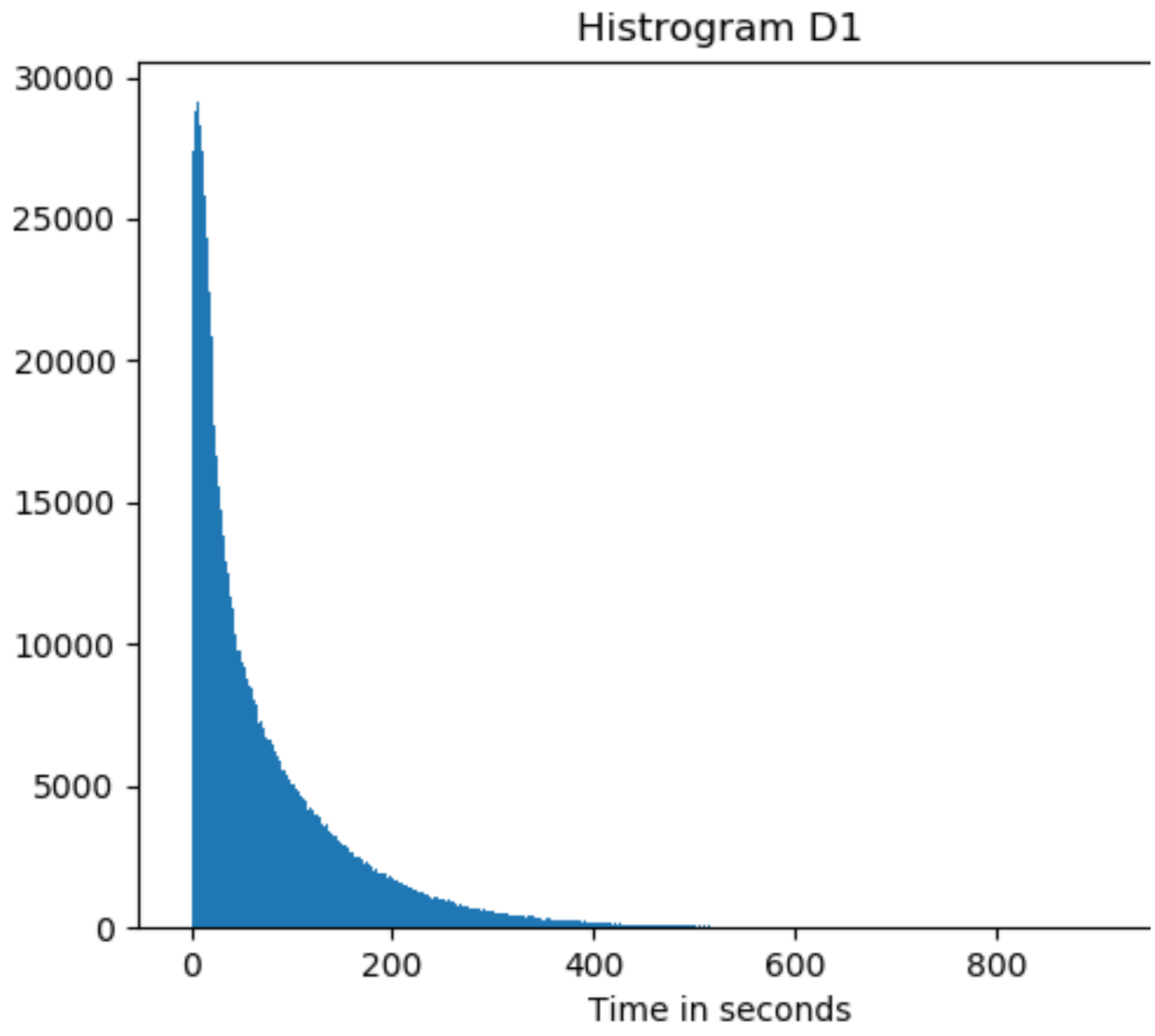
- Vehicles move inside the coverage area only over the rails
- Each rail is seen as a queue
- There are two kind of queues :
  - In: Those are entrances to the coverage area
  - S: Those are the rails which connect to different traffic lights
- Arrivals are following an exponential distribution with mean $1/\lambda$
- Semaphores change between red and green light following a random exponential distribution with mean $1/\gamma$
- Vehicles only move to other place in case of green light and after a random time exponentially distributed with mean $1/\mu$

Movement of the vehicles more similar to real life



Distribution of times inside the coverage area





Why prediction is important

- If the time of permanence in the system is know, it could be possible to know the load of information that needs to be processed

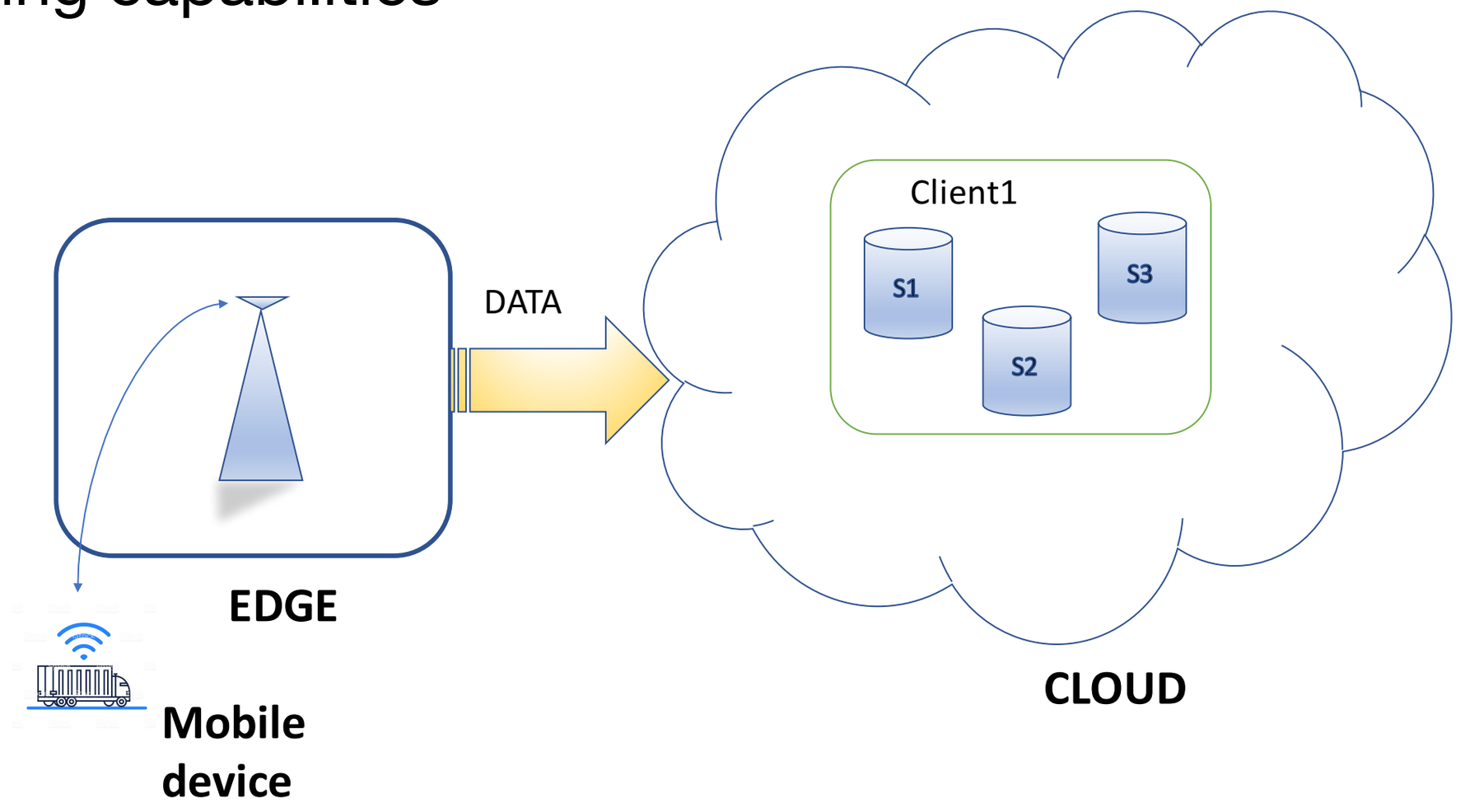
- In order to have a mathematical model of the times inside the system, Coxian distribution is though to be used

- The challenge is find the set of parameters of Coxian distribution that suit the behavior of vehicles

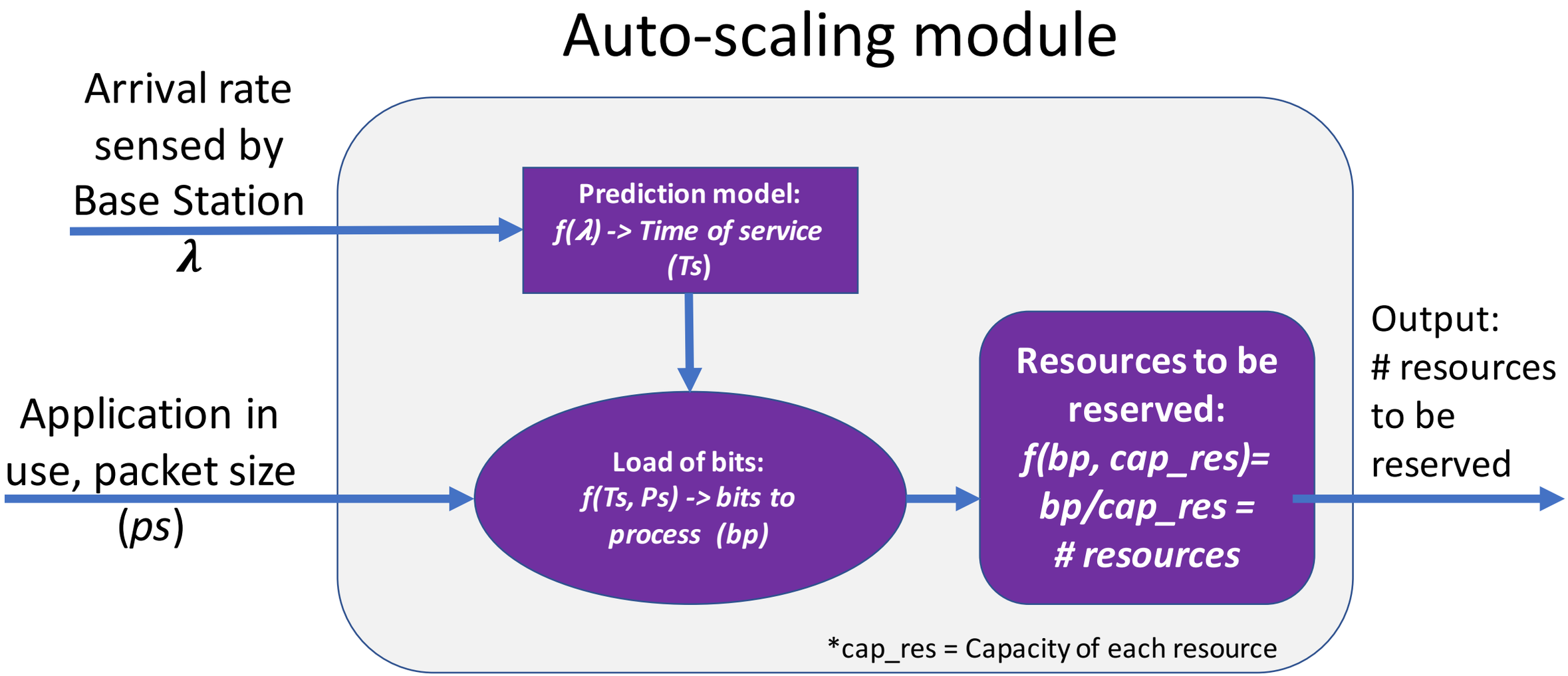
- The model of prediction, is a key point for the Auto-scaling module

System architecture

- Vehicle to infrastructure network model
- Edge processing capabilities
- Edge
  - Prediction
- Cloud
  - Scaling

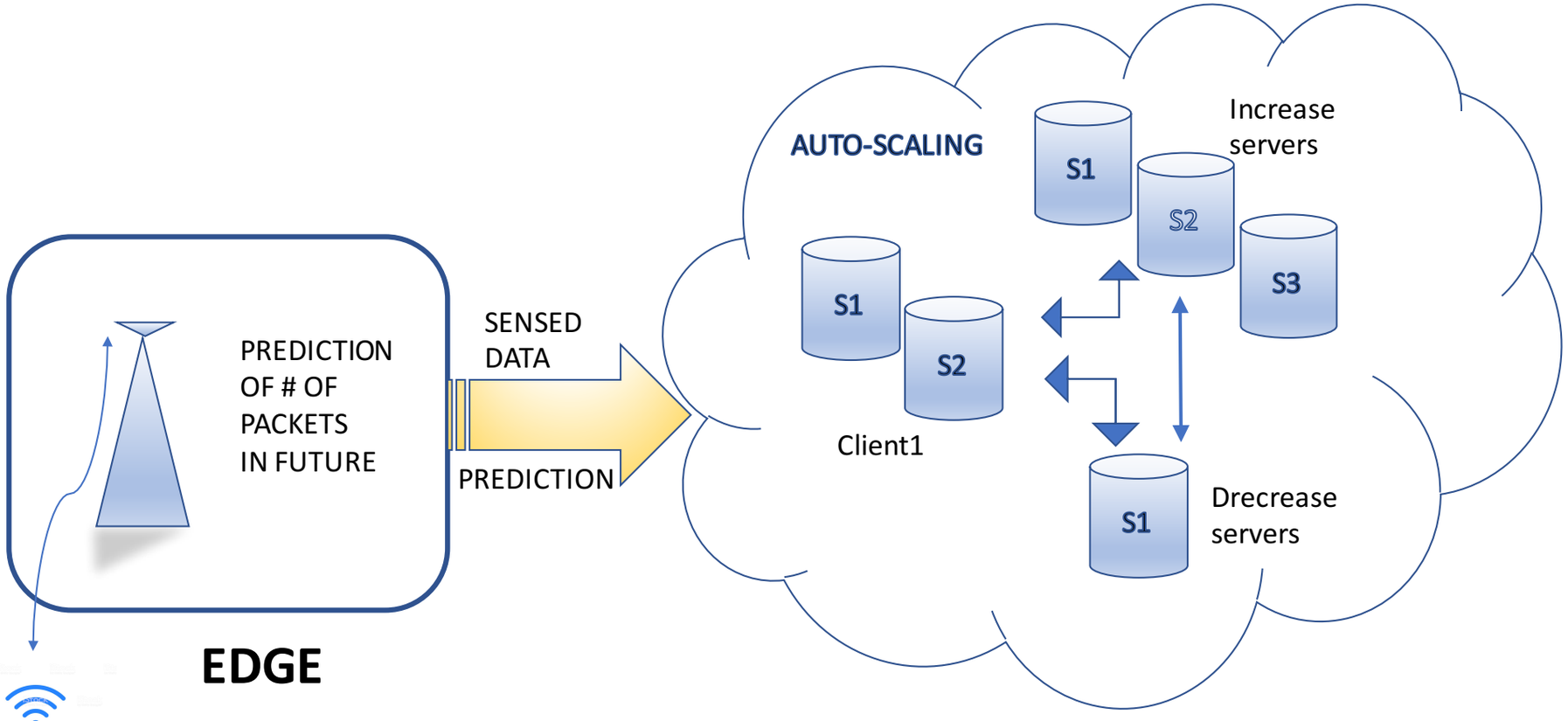


Auto-scaling module



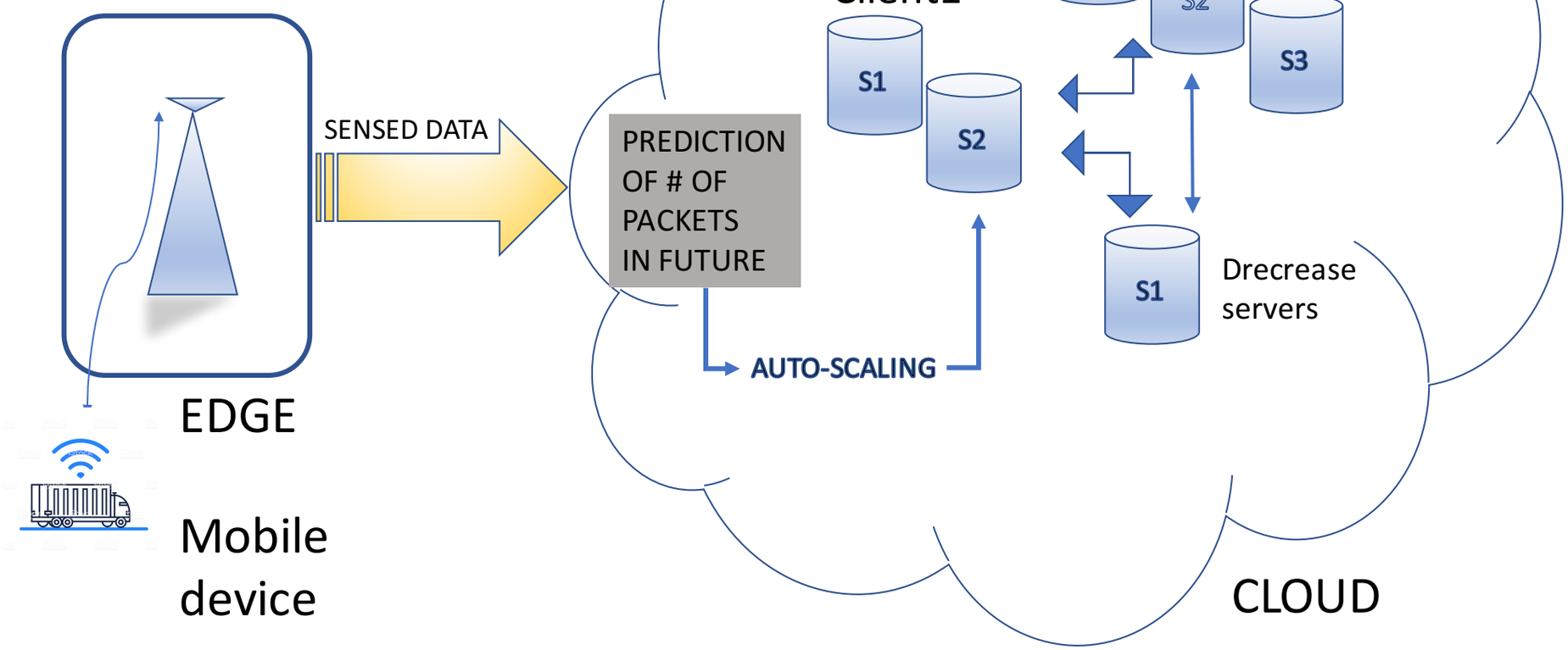
Use of Edge-Cloud architecture

- Two approaches:
  - Prediction made in the Edge, process as much as possible in the Edge and reduce services required from Cloud
  - Prediction made in the Cloud, reserve services in the Cloud



Mobile device

VS



Benefits of Edge-Cloud Auto-scaling

- Knowing the arrival rate to the system, the prediction model estimate the time that vehicles are inside of the system

- Based on the size of packages and the rate of them, (kbps), the load of information is computed

- The work load for a period of time is computed

- Cloud resources are reserved based on the work load

Next steps

- Use of a real Edge-Cloud architecture to be deployed over the Grid5000
- Compare performance of prediction
- Install the prediction module in the Edge and in the Cloud in two separate experiments.
- Compare the accuracy of predictions in Edge against in Cloud

Expected results

- The model is more realistic to real world
- The model could be easily extended to other use-cases
- The prediction made in the Edge is more efficient than in the Cloud
- Latency could be reduced