Abstract—Big cities have to deal with the movement of thousands of vehicles that traverse through the streets every day. As traffic demand increase undesirable problems emerge including traffic congestion, which is becoming more serious day after day. Traffic congestion is an inherently difficult concept resulting to incrementally higher delays, pollution emissions, fuel waste and non-predictable traveling times.

This work introduce a novel architecture that aims to control and manage the utilization of road transport networks aspire to prevent congestion. Inspired by perimeter control approaches this architecture divides an urban area spatially and temporally into smaller regions. The capacity of each road segment within these regions is reserved by users on demand. A Road Site Unit (RSU) is responsible for scheduling vehicle reservations only through congestion free routes. The route reservation problem that suffices is formally derived using mathematical formulations and a real-time algorithm is developed to solve this problem. Preliminary simulation results demonstrate that the proposed architecture offers significant performance improvements in vehicle mobility compared to the current state of operation.

Index Terms—Route reservation architecture, Congestion-free vehicle routing, Intelligent Transportation Systems

I. INTRODUCTION

Transportation research is mainly focused on answering the question of how to eliminate traffic congestion over large scale areas. This is because a large portion of big cities suffers from traffic congestion with sever (in many cases) consequences on personal mobility. Drawbacks of congestion include driver delay and frustration, higher fuel consumption, air pollution and financial losses (in terms of man-hours lost on working days). Congestion has, traditionally, been a difficult problem to tackle since traffic demand fluctuates dynamically. The major cause of congestion is that a portion of the network is reserved to accommodate higher number of vehicles than its actual capacity. Nonetheless, congestion usually occurs due to lack of an efficient management of transport network utilization and not because demand exceeds network’s capacity [1]. Therefore, it is possible to alleviate congestion if vehicles are more effectively distributed over the entire network achieving better load balancing.

An important characteristic of today’s networks is that the majority of drivers prefer to follow main arterial roads instead of using local roads which remain under-utilized most of the time. Besides, drivers usually overload specific portions of the network, creating severe intermittent congestion. Interestingly, undesirable behaviors appear even when real-time traffic state estimates are presented to the drivers. For instance, all rational drivers will prefer to follow less congested roads whenever they become available which could lead to network state oscillations (with congestion continuously shifting between congested to non-congested segments). Moreover, travel time predictions become a very challenging task under this setting, since traffic models can not accommodate driver decisions and as a result the output flow (of congested road segments) becomes extremely sensitive to input demand. Thus, both the road network characteristics and the network state become unpredictable [2].

Considering all the above factors, there is a clear need of better management and more efficient control of the network utilization. Unfortunately, this is a difficult task considering the network size, the volume of vehicles to handle, and the unpredictable driver behavior that changes dynamically according to instantaneous traffic conditions.

Fortunately, an accurate description of the behavior and trends that can be observed in an urban area can be drawn according to the Macroscopic Fundamental Diagram (MFD) [3]. The three parameters of the MFD (i.e., speed, flow and density) are able to fully describe and characterize the state of a network under various traffic conditions. From the MFD, three possible states of road traffic are distinguished: 1) the free-flow state, 2) the saturated state and 3) the congested state (over-saturated state and gridlock) [4].

Observing these three states clearly indicates that as network density increases the possibility of gridlock increase too. If a road segment’s density is maintained bellow critical capacity then the network would be operated in free-flow or saturated state. In the latter case, the possibility of gridlock approaches zero. Hence, a way to avoid congestion is to ensure that the network’s traveling density is maintained bellow critical capacity. This condition ensures that all vehicles are able to travel with free-flow speeds and their travel times can be accurately estimated.

Interestingly, urban areas with multiple centers of congestion observe different characteristics and thus multiple MFDs arise. The existence of multiple MFDs across a region can be detected from the larger scattering present along the MFD [5], [6]. Besides, as introduced by [7], the MFD diagram is well defined if the region is homogeneous (i.e., all road segments have similar characteristics). Hence, to better model the road transport networks, urban areas can be partitioned into a number of small homogeneous regions

This work was partially supported by the European Research Council (ERC) under the ERC Advanced Grant through the FAULT–ADAPTIVE Project.

C. Menelaou, P. Kolios, S. Timotheou, and C.G. Panayiotou are with the KIOS Research Center for Intelligent Systems and Networks, and the Department of Electrical and Computer Engineering, University of Cyprus, {cmene102, pkolios, timotheou.stelios, christosp}@ucy.ac.cy
characterized by a single MFD. The proposed architecture assumes that such clustering is done a priori and each region is identified. Within each region, vehicle routes are controlled by a single road site unit (RSU). All road segments within the same region have the same critical capacity percentage of the region’s MFD.

Instead of calculating and presenting road state estimations to the users, the RSU employs a route reservation algorithm to schedule vehicle routes (both in space and time) through non-congested road segments. To do so, the time horizon is divided into time slots, and for every time slot the RSU keeps track of the number of vehicles that are expected to traverse through each road segment. As vehicles enter each distinct region, their origin-destination pair is send to the RSU which is responsible for scheduling vehicle routes through feasible paths. Feasible paths are those for which road segments do not reach their critical capacity within particular time slots. If segments have reached their critical capacity, these segments become non-available and can not be allocated to any vehicles during those time periods for which road density exceeds its critical capacity. Instead, the RSU can instruct vehicles to wait at their entry point (origin) before starting their journey in the region, or to uses alternative non-congested routes, whenever these alternatives minimize the arrival time to the exit point (destination). Hence, a route reservation is proposed as an alternative approach to state estimation techniques. Given the nominal speed (free-flow) at every road segment, RSUs can approximate the time for each vehicle to traverse each road segment. Therefore, RSUs are able schedule route reservation along consecutive road segment and consecutive time-steps for traversing each region of the network. The objective of the RSU is to schedule vehicles through non-congested road segments while simultaneously minimizing the arrival time at the destination. In this work, it is assumed that all vehicles follow the scheduled paths within the allocated time slots.

In addition to presenting the proposed architecture, the contribution of this work is to develop a real-time scheduling algorithm to solve the route reservation problem. Furthermore, the proposed architecture and the derived route reservation algorithm are analyzed using a realistic road transport scenario in a large area within Nicosia, Cyprus. The rest of the paper is organized as follows. A brief literature review is included in Section II while Section III introduces the proposed architecture. Section IV mathematically describes the route reservation problem and Section V presents a heuristic real-time solution for solving this problem. Preliminaries results included in Section VI demonstrate the benefit of the proposed solution. Finally Section VII concludes this work.

II. RELATED WORK

Perimeter control strategies are the current state of the art approaches trying to control road utilization and to mitigate traffic congestion. These approaches address the congestion problem by first clustering an urban area into homogeneous regions (i.e., regions that can be accurately characterized by a single MFD). Thereafter, various perimeter and boundary flow control strategies are implemented to distribute road traffic within each region as evenly as possible. The main aim of these approaches is to maintain the rate of vehicles that are allowed to enter each region (considered as a “reservoir”) around a desired point, while maximizing the system’s throughput. Within this setting, traffic dynamics are defined using the particular MFD diagram, and no considerations are made on the specific road segment characteristics (i.e., individual densities and flows). Evidently, such kind of approaches greatly reduce the computational burden as discussed in [5] and [7]. The major advantage of the proposed route reservation approach compared to perimeter control is that both inter- and intra-flow control can be achieved in the former case whereas only inflow control can be achieved by perimeter control. At the same time, since the proposed architecture capitalizes on regions of homogeneous behavior, processing at the RSUs can be achieved efficiently in real-time.

Literature indicates that the most of the online approaches that try to control the utilization of road traffic networks employ schedulers that guide vehicles through non-congested regions assuming that the current traffic state is known. Others, base their decision on an estimation of the future network state according to statistical or dynamic traffic assignment models [8] [9]. Importantly, all these approaches focus on minimizing the travel time (instead of the arrival time) by avoiding the utilization of road segments that are on the congested state. However, they fail to take into account the fact that scheduling vehicles on road segments that may be in the saturated state can cause congestion since on saturated state vehicle behavior changes non-linearly with density. Besides, the situation is worsen in the case where the traffic state reaches the over-saturated state and the capacity drop phenomenon occurs. Under this phenomenon, not only is the vehicle speed lower but also the maximum capacity of a road segment drops by up to $15\%$ [10]. Hence, when the travel time of each vehicle is minimized without taking into account the non-linear dynamics of vehicle behavior in congested conditions causes severe inefficiencies. Therefore, there is a need not only to minimize the travel time of each vehicles but, simultaneously, to prevent the network from getting in the over-saturated state that triggers the capacity drop.

The proposed route reservation algorithm is inspired by air traffic control systems that use reservations to solve the ground-holding problem [11]. Today, the amount of air-travelers is steadily increasing while the runway capacity has remained stagnant [12]. As a consequence, airports may (at times) not be able to serve the air traffic demand that rapidly increases. In order to addresses this problem Air Traffic Control and Management Systems (ATC/ATM) decomposed the runway capacity into time slots and each airplane has a specified time slot that must be used for take-off or landing. Compared to road traffic, the air traffic control problem is much simpler since the size of problem is smaller than the road reservation problem. The route reservation problem in transportation networks has been previously proposed in [13]
where an iterative Dijkstra approach (namely, IDA) has been employed to approach this problem. In this paper, an updated version of IDA (namely, IDAU) is investigated. As shown in the sequel, IDAU provides a better solution to IDA but at a higher computational cost. Moreover, for the first time both algorithms are evaluated over a real-world scenario considering a section of Nicosia’s transportation network.

III. PROPOSED ARCHITECTURE

Using today’s advances in information and communication technologies, the proposed architecture offers a complete solution able to manage and control in real-time the utilization of a traffic network. Figure 1 illustrates the basic structure of the proposed architecture where, the key feature is that an urban area is clustered into homogeneous regions based on the properties of a well-defined MFD (Regions 1, 2 and 3). Clustering of these regions can be achieved based on [7] where an urban area is split into homogeneous regions that experience small variances in traffic characteristics (according to the spatial compactness of each region). When a small variance is maintained for the mobility patterns across the different road segments of a region, then accurate and reliable mobility estimates can be obtained (since vehicles would be traveling with predictable mobility patterns). The work in [5] shows that the region-wide mean flow and vehicle density closely approximate the MFD dynamics. At the same time [7] indicates that if a region contains sub-regions with significantly different levels of congestion, the control strategies will be inefficient.

IV. PROBLEM FORMULATION

A homogeneous region is considered as a graph $G = (V, E)$ with vertices $V$ and edges $E$ being the road junctions and segments, respectively. Each road segment $(i,j) \in E$, $\{i,j\} \in V$ has a specified capacity which is obtained from the region’s MFD. The critical capacity of each particular road segment is denoted by $K_{ij}$. The time horizon is discretized into time slots and within each time slot vehicle reservations are made for the use of particular road segments. Evidently, the number of consecutive time slots needed to traverse a particular road segment depends on the road length and the traveling speed. Since reservations ensure that vehicles travel in congestion-free conditions, vehicles can be, safely assumed, to be traveling with their free-flow speed.

In the following section, a systematic approach is followed to derive a mathematical formulation of the route reservation problem that arises within each homogeneous region that an RSU controls.
in time slot \( t \). The reservation state denotes if a road segment has reached its critical density or not. As indicated above, congested conditions can be avoided if all vehicles are traveling through road segments that have not reached their critical capacity. Under non-congested conditions, all vehicles are assumed to travel with free-flow speeds. Let \( \bar{c}_{ij} \) be the time needed for the vehicle to traverse road segment \((i, j)\). Then, \( x_{ij}(t) \) is defined as follows:

\[
x_{ij}(t) = \begin{cases} 
1, & \text{if } r_{ij}(\tau) < K_{ij}, \forall \tau = t, \ldots, t + \bar{c}_{ij} \\
0, & \text{otherwise}
\end{cases}
\]  

where \( x_{ij}(t) = 1 \) denotes the non-congested state, and \( x_{ij}(t) = 0 \) the congested state.

Route reservations are allowed only along road segments and during time slots where \( x_{ij}(t) = 1 \). Alternatively, a vehicle should wait at its origin (node \( o \)) until a non-congested path becomes available, or be rerouted through an alternative road segments that is congestion-free. Clearly, a combination of the two aforementioned options can be explored, i.e., wait for a short period at node \( o \) and then take an alternative path. The decision is made solely based on which solution achieves the earliest arrival time to the destination.

Considering the above notation, the cost of traversing a road segment \( c_{ij}(t) \) can be expressed as follows:

\[
c_{ij}(t) = \begin{cases} 
\bar{c}_{ij}, & \text{if } x_{ij}(t) = 1 \\
\infty, & \text{if } x_{ij}(t) = 0 \text{ and } i \neq s \\
\bar{c}_{ij} + W_o(t), & \text{if } x_{ij}(t) = 0 \text{ and } i = s
\end{cases}
\]  

where, \( W_o(t) \) denotes the smallest number of time slots that a vehicle should wait at \( o \) such that a feasible path is found to traverse from \( o \) to destination \( e \) through non-congested road segments.

**Earliest Destination Arrival Time problem (EAT):**

Given an origin-destination pair of a vehicle, the traversal start time \( t_0 \), and the cumulative reservation state \( x_{ij}(t) \), then the EAT problem requires the earliest-arrival-time path from \( o \) to \( e \) starting at \( t_0 \). Let \( p_k \) denote the \( k \)-th path from source \( o \) to destination \( e \). Then an arbitrary path can be expressed as \( p_k = (v_1, v_2, v_3, \ldots, v_{nk}) \), where \( v_j \in V \) is the \( j \)-th visited node in the \( k \)-th path, with \( v_1 = s \) and \( v_{nk} = e \) and \( n_k \) is the length of the path. Also, let \( \tau_j(w) \) denote the earliest arrival time at junction \( v_j \) assuming that the vehicle waits \( W_o(t) \) time steps at the origin before commencing its journey. Then, the earliest arrival time to each node of the path can be expressed as follows:

\[
\begin{align*}
d_{0}(t_0) &= t_0 + W_o(t), \quad W_o(t) \geq 0 \\
n-k(t_0) &= d_{0}(t_0) + c_{v_k,v_k} (d_{v_k}(t_0)) \\
&\vdots \\
n-k(t_0) &= d_{n_k}(t_0) + c_{v_{n_k},v_{n_k}} (d_{v_{n_k}}(t_0)) 
\end{align*}
\]  

Overall, the EAT problem can be expressed in compact form as follows:

\[
EAT = \min_{w, p_k} d_{n_k}(t_0) 
\]  

**V. Iderative Dijkstra Algorithm Updated (IDAU)**

IDAU is an improved version of the IDA algorithm already proposed in [13]. Both algorithms are based on the well-known Dijkstra shortest path solution (which uses the label-setting property and the relaxation technique) that is modified in order to calculated the earliest-destination-arrival-time on each road junction (i.e., \( d_i = \min(d_i, d_j + c_{ij}(t)) \)) [14]. Since the EAT problem is shown to be an NP-complete problem in [15], IDA and IDAU are both heuristic solutions.

Similarly to IDA, IDAU solves EAT in two stages. The inner stage returns the earliest-destination-arrival-time path by allowing delay at the beginning of each road junction. This is a lower bound solution, which can be solved optimally, since this sub-problem can be classified as a shortest path problem with dynamically changing weights. In those cases when the accumulate reservations of a particular road segment reach critical capacity (i.e., \( x_{ij}(t) = 0 \)), then a vehicle is not allowed to traverse the region and should wait until feasible reservation schedules can be made. Noticeably, with each iteration of the inner stage, the weights of each road segment change dynamically according to the waiting intervals (i.e., \( w_{ij}(t) \cap (i, j) \in E \)) that a vehicles should incur while traveling along congestion-free road legs. In total, the traversal time for each road segment includes the travel time (assuming free-flow conditions) and the waiting time at intervals that may need to take place (i.e., \( c_{ij} = \bar{c}_{ij} + w_{ij}(t) \)).

The outer stage is responsible of examining the solution returned by the inner stage and assess if it involves waiting at any of the intermediate road junctions. When no waiting is necessary at intermediate road junctions or when there is only waiting at the entry point into the region then the reservation is granted. On the contrary, when waiting is necessary at any intermediate junction, the outer stage transfers the minimum interval observed along the path at the origin (i.e., the entry point to the region) and updates the traversal start time (i.e., \( t_0 = t_0 + \min(w_{ij}(t)) \)). Thereafter, the inner stage reiterates with the updated start time \( t_0 \). This procedure repeats until no further waiting is required at intermediate junctions.

The major difference between IDA and IDAU is that IDA shifts the accumulated waiting time (obtained from the earliest-destination-arrival-time path) to the origin at the outer stage while the IDAU transfers only the minimum waiting time to the origin when such waiting is necessary.

Hence, as input demand increases the inner stage of IDAU reiterates much more times than that of the IDA since smaller incremental time shifts are made. Nevertheless, both algorithms have the same complexity in the worst case scenario since the number of iterations is bounded by the path length (in case where all road segments have non-congested states). The complexity of both algorithms (using binary heap structures) is \( O(E^2 \log(V)) \). In practice, the
IDAU needs much more iterations to converge and in that sense its execution time is slower than that observed by the IDA. However, the IDAU algorithm has been developed to discover if a more exhaustive solution can offer better route reservations and improve vehicle travel times.

**VI. SIMULATION SETUP AND RESULTS**

One of the major purposes of this work is to examine the improvement that can be achieved if a route reservation algorithm (IDAU) is applied to a single region. A realistic network setup has been considered, with the actual road network topology of a larger area in Nicosia, Cyprus extracted from Open Street Map (http://www.openstreetmap.org) as shown in Figure 3. The selected network contains three major arterial roads (namely, 'Keneti’, ‘Makariou’ and ‘Spirou Kiprianou’ avenues) that create congestion conditions during rush hour. This area consists of 610 road segments and 253 road junctions.

The SUMO mobility simulator [16] was employed to create realistic mobility patterns according to the Krauss car following model [17]. Moreover the TraCI interface [18] was used to control and manage vehicle interactions through the OMNET++ network simulator. Both, the IDAU and the IDA algorithms were coded in OMNET++ to compute the route reservation schedules. The results obtained from each of the two solutions were compared against the traditional behavior (TB) experienced by vehicles when no reservations are made. For the evaluation of the examined algorithms, Monte-Carlo simulations were conducted for different flow-rates (500 – 8000 veh/h), executing a total of 10 simulation for each case. Vehicle arrivals were modeled using the Poisson distribution and random origin-destination pairs were assigned. In the results depicted below, all vehicles were assumed to follow the reserved paths at all times and an ideal communication channel is assumed to exist across vehicles and the RSU. The car-following model parameters are set as follows: vehicle length 5m, maximum speed 15m/s, acceleration 2.5m/s², deceleration 4.5m/s² and the minimum gap distance is 2.5m. Simulation time was set to 1 hour and only vehicles that completed their journeys were considered in the our results. Finally, no overtaking was allowed to ensure that all vehicles followed FIFO queuing (i.e., $t + c_{ij}(t) \leq (t+1) + c_{ij}(t+1), \forall (i,j) \in E$).

Fig. 4 illustrates the average number of vehicles that completed their journeys as a function of the different flow rates. The dashed lines represent the average number of vehicles that have reached their destinations within the simulation time and the scattered plots are the realizations that are obtained by each simulation execution.

![Fig. 4: Number of vehicles that completed their journeys.](image)

Likewise, the dashed lines in Fig. 5 illustrate the average travel time for the different flow rates simulated while the scatter plots show the results obtained for the 10 realizations.

![Fig. 5: Vehicle travel times under varying flow rates.](image)

Fig. 4 and Fig. 5 clearly demonstrate that by applying both the IDA and IDAU reservation algorithms the network performance is greatly improved compared to TB, especially when demand is high. For instance, at a flow rate of 7000 veh/h, less than 3000 vehicles manage to complete their journeys while IDA and IDAU enable more than 5000 vehicles to travel through the region. At the same time, the travel time is greatly reduced. Using IDA and IDAU the mean travel time is only 150s while for TB the mean travel time is more than 400s. It should be noted here that at flow rates of 8000 veh/h, there are instances where TB achieves...
small travel times as well. This is due to the fact that only very few vehicles manage to reach their destination and they do so by following paths at the outskirt of the network avoing congestion that occurs along the main arterial roads. Evidently, at low flow rate, IDA and IDAU have similar behavior to that of TB since no congestion is experienced.

Further, Fig. 6 illustrates the travel time distribution for the flow rate of 8000veh/hour where only vehicles that had reach their destination are shown. The travel time distribution indicates the improvement that can be achieved in the case that a route reservation method is applied instead of TB. This observation clearly motivates the application of such kind of algorithms in real-world solutions. In numbers, the mean travel time of IDA is 121 s, while that of IDA is 130 s and for TB is 327 s. The standard deviation of IDAU is 56, for IDA is 87 and for TB is 425 seeing that as demand increase the travel time of IDA and TB increases at a higher rate than that of IDAU. In either case, these results indicate the robustness of the IDAU and IDA algorithms compared to TB.

As depicted in Fig. 6, IDAU achieves better performance compared to IDA. However, IDAU is considerably more resource demanding than IDA. For example, to obtain the results in Fig. 6, 42 simulation hours were needed instead of just 25 simulations hours for IDA. Therefore, the marginal increase in utilization (and the respective travel time improvements) achieved by IDAU may not be achieved in real-time under practical considerations. Nevertheless, IDAU demonstrates the potential improvements that could be achieved compared to IDA if and when the need arises. In either case, IDA and IDAU provide considerably better road utilization and greatly improve vehicle travel times especially during rush hours.

VII. Conclusions

This work introduced a novel architecture design for road transportation management and control. It also illustrated the potential benefit of applying route reservation strategies over small homogeneous regions under the supervision of a single RSU. Simulation results indicates that the application of a route reservation algorithm over a specific region can ensure congestion-free operation and great improvements in road utilization and vehicle travel times. The updated version of IDA algorithm was shown to increase road utilization compared to the basic IDA, at the expense however of higher execution times.

Future challenges include the design and simulation of multi-region scenarios where RSUs would be cooperating to improve inter-region utilization and maintain a congestion-free operation across larger parts of the network.

References