IoT-Based Crowd Management Framework for Departure Control and Navigation

Ahmed Elbery, Member, IEEE, Hossam S. Hassanein, Fellow, IEEE, Nizar Zorba, Senior Member, IEEE, and Hesham A. Rakha, Fellow, IEEE

Abstract—This paper exploits crowdsensing to propose a novel IoT-Based Vehicle Crowd Management (IoT-VCM) framework. By efficiently managing vehicle departures and navigation, the IoT-VCM clears the network in a shorter time, while maintaining the network at low congestion levels to reduce the average travel time. To compromise between these conflicting objectives, the proposed system encompasses two subsystems that work in harmony, namely; the Travel-Time System-Optimum Navigation (TTSON) and the Vehicle Departure Control (VDC). The IoT-VCM uses different network sensory devices (connected vehicles and smartphones) to collect network information that is fused to compute the current road state conditions, based on which, the VDC determines the allowable vehicle departure rates, and the TTSON optimizes their navigation. The proposed system is developed in a microscopic traffic simulator and tested on a calibrated simulated real network. The IoT-VCM controller is compared to the state-of-the-art techniques reported in the literature, namely the dynamic time-dependent incremental user-optimum traffic assignment.

Index Terms—Crowd management, IoT, system optimum navigation, vehicle departure control, stochastic routing, constrained routing.

I. INTRODUCTION

In recent years, vehicular crowds have become prevalent in urban areas. Sports events at stadiums are typical cases of vehicular crowds where a large number of attendees (that may exceed 100 000) gather in a small area. A large portion of those attendees drive to these events, which means tens of thousands of vehicles need to leave the area after the event. Such vehicular crowd will cause severe congestion if they are not smartly and efficiently controlled in both departure sequencing and traffic assignment. In addition to sports events, other types of social events, such as music festivals or concerts, also draw large vehicular crowds. Another case for vehicular crowds is an emergency evacuation due to natural disasters [1].

The high vehicular traffic volume in such cases exceeds the road network capacity. Cities are not designed to support such high traffic volumes for many reasons. First of all, these events have non-recurrent or occasionally occur infrequently, e.g., every couple of months or years. Secondly, the high cost of increasing the capacity of the current road networks [2]. This cost is not only financial but also administrative because there are other types of networks (i.e., water networks) intersecting with the road network; consequently, many other authorities would have to be involved. Thirdly, in some cases, it is not possible to increase the road capacity because of the area limitation or infrastructure limitations (e.g., bridges). Lastly, such events need special traffic control techniques [3], which may include customized traffic signal timing combined with blocking some roads and detouring others.

By coupling this high traffic volume and the road network capacity constraints, one can soon realize how vehicular crowds can adversely affect mobility in such cases. From a driver’s perspective, the high traffic volume increases the vehicle density on the roads and results in lower average travelling speeds and longer travel times, according to the fundamental relationship between speed and density [4] shown in Fig. 1. From the road link level perspective, increasing the vehicle density above a predefined threshold, known as density-at-capacity, denoted as $K_c$, in Fig. 1, moves the network towards the congested regime, resulting in lower flow rates and lower vehicular speeds. Thus, the best link utilization is achieved at a vehicle density of $K_c$, the jam density, which produces the maximum flow on the link $Q_{Max}$ shown in Fig. 1. $Q_{Max}$ is known as the link capacity, which is the maximum number of vehicles that can exit a road segment per unit time [5]. In Fig. 1 also, we can see the non-linear relationships between the three parameters, which make traffic assignment a non-trivial problem, especially in large networks with different road segments and different parameters for each segment.

Fig. 1. Fundamental relationships between vehicle density, speed, and flow rates.
It is obvious that higher traffic demand increases vehicle queuing at intersections, which increases travel time. Moreover, in some cases, vehicles may stop at traffic signals for more than one cycle, and in complex road networks, this may result in network grid-locks.

Following large events, the high traffic demand and its impact on the mobility in such constrained road networks bring forward the need for efficient and smart traffic control and route planning techniques. Such techniques should efficiently utilize the network facilities while maintaining a certain performance level that satisfies the drivers’ expectations.

The widespread proliferation of the Internet of Things (IoT) [6], including transportation applications [7], [8], opens new doors to boost mobility and reduce congestion problems in smart cities. The ubiquitous use of smartphones and the foreseeable prevalence of connected vehicles pave the way to building new IoT paradigms that exploit crowdsensing to tackle city-wide complex problems.

In this paper, we utilize crowdsensing and IoT to introduce a novel vehicular crowd management framework, the IoT-VCM, that can better manage vehicle crowds after special events or in emergency evacuation cases. The proposed framework takes advantage of crowdsensing to collect data from drivers through smartphones or connected vehicles. This information is used to compute the network state conditions, that are used to optimize the route assignment and manage vehicle departures.

The objectives of the proposed system are to clear the network faster and, in the meantime, maintaining low average travel times, by limiting the vehicle density on each road segment to its density-at-capacity $K_c$. Achieving these objectives is challenging because of the trade-offs between them. In order to clear the network faster, there are two requirements. Firstly, vehicles are required to depart as early as possible. Secondly, vehicles should travel at high speed to minimize travel time. However, these two requirements conflict with each other because allowing high vehicle departure rates results in high traffic volumes, which increases the congestion level, consequently, increases the travel time. On the other hand, by reducing the vehicle departure rate, vehicles will travel at higher speed and experience shorter travel times, but some vehicles will start their trips very late, which will again increase the network clearance time.

To address this problem, the proposed system offers a compromise between these two objectives (the maximum allowable departure rates and lower travel times) by integrating two subsystems:

1) **The Vehicle Departure Control (VDC)** which is responsible for controlling vehicle departures to allow vehicles to leave as early as possible while maintaining a maximum congestion level. It achieves this target by computing the maximum allowable vehicle departure rates based on both the current states of road links and their capacities.

2) **The Travel-Time System Optimum Navigation (TTSON)** which minimizes the network-wide travel time. TTSON stochastically assigns routes to vehicles in such a way that utilizes the available alternative routes, and at the same time, considers their available capacities in order to avoid congesting these roads.

The IoT-VCM uses the collected data to compute the road state conditions (traffic volume and travel time on each road segment) in real-time. The road state condition information is used along with historical data by the TTSON to optimize the network-wide vehicle route assignment using Linear Programming (LP) [9]. In the formulated LP optimization problem, the link capacities and their current traffic loads are used to constrain the network congestion.

The same optimization model is used by the VDC to compute the maximum allowable vehicle departure rates after the event, hence, the vehicle inter-departure intervals. Periodically, the VDC notifies the drivers about their estimated departure times. In the case of high traffic demand, some drivers may be requested to postpone their departure. As a motivation, the system can offer them direct incentives such as discounted tickets for other events, or spending some time to take pictures and receive signatures from public figures. A more advanced incentive mechanism can be utilized such as the social-aware incentive mechanism based on deep reinforcement learning proposed in [10].

The proposed system is developed within a microscopic simulator and tested on a real road network with real calibrated traffic. The 2022 FIFA World Cup, which will be held in Doha, Qatar, is used as the case study. The Doha road network is implemented and used to compare our proposed system to the dynamic time-dependent incremental user-optimum traffic assignment as a typical real-time navigation system that is currently in use (e.g., Waze and Google Maps).

The remainder of the paper is organized as follows. Section II explores the related work. Section III explains the network model. Section IV describes the IoT-VCM system and its components. Section V tackles our case study on the Doha network in Qatar and the results. The final conclusions are presented in Section VI.

## II. RELATED WORKS

The advances of vehicular communication and IoT have promoted smart mobility and navigation systems that use vehicles and mobile devices as sensors to collect real-time information. Online services, such as Google Maps and Waze, provide online dynamic navigation guidance based on the estimated travel time information collected from vehicles or mobile devices. Other online services, such as INRIX [11], provide real-time traffic information to assist drivers and autonomous vehicles selecting routes. Some research efforts [12], [13] propose a mobile crowdsensing that uses both current and historical information to predict traffic conditions and travelling speeds to enable the dynamic routing of drivers wishing to avoid congestion. All these guidance systems typically do an all-or-nothing traffic routing by assigning all vehicles to the shortest path (which are user equilibrium models) or, at most, provide alternative routes without considering the system-wide performance or trying to minimize network-wide congestion.

Other research efforts consider balancing the traffic across alternative routes. For instance, [14] proposes a heuristic approach...
to randomly assign vehicles to different routes based on each vehicle’s remaining travel time and the route popularity. Compared to this algorithm, our proposed model is a step up because we utilize optimization to assign routes instead of assigning routes randomly. Moreover, our proposed model accounts for the network-wide state conditions in assigning the traffic (i.e., the current traffic load on each road segment).

An evacuation route planning model presented in [15] computes the relationship between clearance time, number of evacuation paths, and congestion probability during an evacuation. The model in [15] does not utilize vehicular networks and does not account for the current network conditions. Instead, it focuses on road capacity uncertainty. In our work, we collect actual road information, from which we can estimate the available road capacities. Consequently, a portion of the traffic can be assigned to underutilized routes. In [16], the authors develop a system-optimum eco-routing model to minimize the total system-wide fuel consumption. By utilizing multiple routes, they are able to reduce traffic congestion and fuel consumption. Compared to [16], our proposed model focuses on the objective of clearing the network faster by enabling vehicles to use alternative routes and by adjusting the vehicle departure times based on the network conditions. The authors in [17] utilize crowdsensing to propose a personalized route planning mechanism based on the quality of the road surface. However, they use shortest path techniques and do not consider the system-wide performance. The authors in [18] propose a grid road network model for path planning in emergency situations. They use a shortest path technique which is applied to the grid road network model they develop.

In literature, crowd management and crowdsensing have been used in different ways to serve different purposes. For instance, in [19], the authors propose a framework to resolve congestion by controlling the movement of crowds of persons. In [20], [21], the authors study the task assignment cost and performance in mobile crowdsensing, and based on that, they combined the path planning and the task assignment using different techniques to improved task assignment performance.

To the best of our knowledge, none of those previous works on crowd management use vehicle departure control as being proposed in this paper.

III. NETWORK MODEL AND PROBLEM DEFINITION

In the proposed system, the road network is represented by a directed graph $G(N, L, Q)$, as shown in Fig. 2. In this graph, $N = \{i : i = 1, 2, \ldots, n\}$ is a set of $n$ nodes (e.g., road intersections), and $L$ is a set of $l$ directed links, each link is a road segment, i.e., $L = \{L_{i,j} : i, j \in N\}$, where $L_{i,j}$ is the road segment from node $i$ to node $j$. Each road segment has a capacity $Q_{i,j}$, which is the maximum traffic flow rate that can exit (or enter) this link [22], [23] (corresponding to $Q_{\text{Max}}$ in Fig. 1). Moreover, each segment has a variable travel time $T_{i,j}$ that depends on the traffic conditions on the segment. Additionally, each road link has a time-varying traffic load $\zeta_{i,j}$ that represents the average vehicle rate passing it. Thus, the available capacity

$$\hat{Q}_{i,j}$$

on the road segment $L_{i,j}$ is the difference between its capacity and its current load, i.e., $\hat{Q}_{i,j} = Q_{i,j} - \zeta_{i,j}$.

The vehicular crowd in the network is represented by the red area in Fig. 2. In our case study, this red area is a stadium where some soccer matches will be held. The destinations of vehicles are distributed across the network. Consequently, vehicles in this crowd are grouped into a set of traffic flows, known as Origin-Destination (OD) demand pairs [24]. These traffic flows are built based on the driver destination information collected from the smartphones as described in Section IV-B. Each flow is a set of vehicles that will leave the crowded area after the event to a given destination.

There is a set $F$ of $f$ concurrent flows, each flow is identified by $k \in \{1, 2, \ldots, f\}$, and each has $V_k$ cars that need to leave within a time interval $\tau$. Thus, each flow $k \in F$ has a traffic rate $q_k$ (in vehicles per hour $\text{veh}/\text{h}$), where $q_k = V_k/\tau$. The initial value for the time interval $\tau$ is calculated based on the information collected from the drivers. We assume these vehicles are connected and follow the route recommendations they receive from the Traffic Management Center (TMC). We call these flows the Controlled Traffic (CT).

In addition to the CT, there is Background Traffic (BT), which is the regular day-to-day traffic traversing the network. Vehicles in the BT use the dynamic time-dependent incremental user-optimum (Nash optimum) traffic assignment.

Using crowdsensing, data processing, and optimization techniques, we aim to clear the network in a shorter time using two techniques. The first is the stochastic system optimum routing, which is the core function of the TTSON subsystem. Its goal is to efficiently utilize the network resources by simultaneously routing CT vehicles (going to the same destination) through multiple alternative routes, in such a way that minimizes the network-wide travel time (system-optimum traffic assignment). The second technique is the departure control that computes the CT vehicle maximum allowable departure rates and controls vehicle departure times based on these rates to avoid network congestion. These two techniques are jointly optimized to allow vehicles to depart as early as possible while avoiding network congestion. This coupling results in reasonably high
average speeds. Consequently, it clears the network in a shorter time.

IV. IOT-BASED VEHICLE CROWD MANAGEMENT FRAMEWORK

The IoT enables the integration of several technologies and communications paradigms such as identification and tracking technologies, wired and wireless sensor and actuator networks, communication protocols, and smart devices [25]. Our proposed system utilizes IoT by integrating different sensors (connected vehicles, smartphones) and actuators (autonomous vehicles and human drivers) to enable better crowd management techniques for vehicular crowds. We will refer to “connected vehicle” to both smart vehicles or non-smart vehicles with a passenger holding a connected smartphone.

This section describes the proposed framework and its components, including the data collection and processing, the optimization process, the stochastic navigation subsystem, and the departure control subsystem.

The front-end of the system is an application that can be installed on vehicle on-board computers or smartphones. Using the Global Positioning System (GPS) capabilities in these devices, this application computes the travel time that the vehicle experiences on each road segment. Then, it communicates this information to the TMC. Additionally, it can request routes from the VDC. Through this interface, the user can set the destination information to the TMC. Additionally, it can request routes from drivers going to the same destination area are aggregated in the same traffic flow. The number of vehicles in each flow is used to estimate the flow’s initial average traffic rates based on exponential distribution headway intervals. These rates will be used as initial rates by the optimization module and can be updated by the VDC based on the network conditions.

1) Travel Destinations and Flows: The destination information from drivers is necessary for estimating the traffic flows and their distribution across the network. During or before the event, the system polls the destination from divers. The driver destinations are used to build the traffic flows, where the system divides the city into a set of destination areas. All the drivers going to the same destination area are aggregated in the same traffic flow. The number of vehicles in each flow is used to estimate the flow’s initial average traffic rates based on exponential distribution headway intervals. These rates will be used as initial rates by the optimization module and can be updated by the VDC based on the network conditions.

2) Traffic Information From Vehicles: Data collected from connected vehicles are used, along with other data sources (such as road network information, i.e., free-flow speeds, capacities, and lengths) to continuously update the network state conditions, including link travel times $T_{i,j}$ and traffic loads $\zeta_{i,j}$.

In order to compute these parameters, we utilize the GPS in connected vehicles or smartphones. Through the GPS, the front-end application can identify which link it is traversing. Whenever a vehicle enters a new road segment, the application initializes the travel time to the current time, and when the vehicle exits the road segment, it computes the link travel time $T_{i,j}$. Subsequently, the application builds a message containing this information along with the current time. The application then tries to send this message to the TMC using either V2I (such as car A in Fig. 3) or V2V (such as car B in Fig. 3). If the car does not have a connection (such as car C in Fig. 3), it will store the message until reaching an area covered by the network, or finding a vehicle-to-vehicle path to the TMC.

The TMC drops the expired messages based on the message timestamp. Upon receiving an unexpired message, it uses the received travel time to update the link information. The received link travel times may be noisy due to different driver behaviours vehicle, it performs the required processing such as authentication, decryption, and unpacking. Then, based on the message contents, it forwards it to the appropriate module. For instance, route request messages are forwarded to the navigation module. Through the communication module, all the components can also send messages to the users through the communication network which is the infrastructure that communicates the system back end to different types of sensors/actuators. The communication depends on the types and capabilities of the network and the different devices. For instance, smartphones can use cellular or WiFi networks. Vehicles can use their VANET [26] or the 5G [27] communication capabilities using either vehicle-to-infrastructure (V2I) or Vehicle-to-Vehicle (V2V) multi-hop connections [28].

B. Data Collection and Processing

To optimize navigation and vehicle departures, the proposed system needs to compute two parameters for each road segment $L_{i,j}$: its travel time $T_{i,j}$ and its current traffic load $\zeta_{i,j}$. Moreover, it needs to estimate the traffic flows and their rates. The system employs crowdsensing to collect data from different sources to infer these network state conditions.

The communication module receives a message from a connected device, this application computes the travel time that the vehicle experiences on each road segment. Then, it communicates this information to the TMC. Additionally, it can request routes from the TMC. Through this interface, the user can set the destination information to the TMC. Additionally, it can request routes from drivers going to the same destination area are aggregated in the same traffic flow. The number of vehicles in each flow is used to estimate the flow’s initial average traffic rates based on exponential distribution headway intervals. These rates will be used as initial rates by the optimization module and can be updated by the VDC based on the network conditions.

The back-end system is the TMC, which has the following modules; data collection and processing module, optimization module, VDC, and TTSON. The system architecture is shown in Fig 3.

A. Communication Module

As shown in Fig. 3, the communication module is responsible for communicating the different system modules to external components such as vehicles or smartphones. When the communication module receives a message from a connected vehicle, it performs the required processing such as authentication, decryption, and unpacking. Then, based on the message contents, it forwards it to the appropriate module. For instance, route request messages are forwarded to the navigation module. Through the communication module, all the components can also send messages to the users through the communication network which is the infrastructure that communicates the system back end to different types of sensors/actuators. The communication depends on the types and capabilities of the network and the different devices. For instance, smartphones can use cellular or WiFi networks. Vehicles can use their VANET [26] or the 5G [27] communication capabilities using either vehicle-to-infrastructure (V2I) or Vehicle-to-Vehicle (V2V) multi-hop connections [28].

The destination information from drivers is necessary for estimating the traffic flows and their distribution across the network. During or before the event, the system polls the destination from divers. The driver destinations are used to build the traffic flows, where the system divides the city into a set of destination areas. All the drivers going to the same destination area are aggregated in the same traffic flow. The number of vehicles in each flow is used to estimate the flow’s initial average traffic rates based on exponential distribution headway intervals. These rates will be used as initial rates by the optimization module and can be updated by the VDC based on the network conditions.

The communication module receives a message from a connected device, this application computes the travel time that the vehicle experiences on each road segment. Then, it communicates this information to the TMC. Additionally, it can request routes from the TMC. Through this interface, the user can set the destination information to the TMC. Additionally, it can request routes from drivers going to the same destination area are aggregated in the same traffic flow. The number of vehicles in each flow is used to estimate the flow’s initial average traffic rates based on exponential distribution headway intervals. These rates will be used as initial rates by the optimization module and can be updated by the VDC based on the network conditions.

The back-end system is the TMC, which has the following modules; data collection and processing module, optimization module, VDC, and TTSON. The system architecture is shown in Fig 3.

The destination information from drivers is necessary for estimating the traffic flows and their distribution across the network. During or before the event, the system polls the destination from divers. The driver destinations are used to build the traffic flows, where the system divides the city into a set of destination areas. All the drivers going to the same destination area are aggregated in the same traffic flow. The number of vehicles in each flow is used to estimate the flow’s initial average traffic rates based on exponential distribution headway intervals. These rates will be used as initial rates by the optimization module and can be updated by the VDC based on the network conditions.

The communication module receives a message from a connected device, this application computes the travel time that the vehicle experiences on each road segment. Then, it communicates this information to the TMC. Additionally, it can request routes from the TMC. Through this interface, the user can set the destination information to the TMC. Additionally, it can request routes from drivers going to the same destination area are aggregated in the same traffic flow. The number of vehicles in each flow is used to estimate the flow’s initial average traffic rates based on exponential distribution headway intervals. These rates will be used as initial rates by the optimization module and can be updated by the VDC based on the network conditions.

The back-end system is the TMC, which has the following modules; data collection and processing module, optimization module, VDC, and TTSON. The system architecture is shown in Fig 3.

The destination information from drivers is necessary for estimating the traffic flows and their distribution across the network. During or before the event, the system polls the destination from divers. The driver destinations are used to build the traffic flows, where the system divides the city into a set of destination areas. All the drivers going to the same destination area are aggregated in the same traffic flow. The number of vehicles in each flow is used to estimate the flow’s initial average traffic rates based on exponential distribution headway intervals. These rates will be used as initial rates by the optimization module and can be updated by the VDC based on the network conditions.

The communication module receives a message from a connected device, this application computes the travel time that the vehicle experiences on each road segment. Then, it communicates this information to the TMC. Additionally, it can request routes from the TMC. Through this interface, the user can set the destination information to the TMC. Additionally, it can request routes from drivers going to the same destination area are aggregated in the same traffic flow. The number of vehicles in each flow is used to estimate the flow’s initial average traffic rates based on exponential distribution headway intervals. These rates will be used as initial rates by the optimization module and can be updated by the VDC based on the network conditions.

The back-end system is the TMC, which has the following modules; data collection and processing module, optimization module, VDC, and TTSON. The system architecture is shown in Fig 3.

The destination information from drivers is necessary for estimating the traffic flows and their distribution across the network. During or before the event, the system polls the destination from divers. The driver destinations are used to build the traffic flows, where the system divides the city into a set of destination areas. All the drivers going to the same destination area are aggregated in the same traffic flow. The number of vehicles in each flow is used to estimate the flow’s initial average traffic rates based on exponential distribution headway intervals. These rates will be used as initial rates by the optimization module and can be updated by the VDC based on the network conditions.

The communication module receives a message from a connected device, this application computes the travel time that the vehicle experiences on each road segment. Then, it communicates this information to the TMC. Additionally, it can request routes from the TMC. Through this interface, the user can set the destination information to the TMC. Additionally, it can request routes from drivers going to the same destination area are aggregated in the same traffic flow. The number of vehicles in each flow is used to estimate the flow’s initial average traffic rates based on exponential distribution headway intervals. These rates will be used as initial rates by the optimization module and can be updated by the VDC based on the network conditions.

The back-end system is the TMC, which has the following modules; data collection and processing module, optimization module, VDC, and TTSON. The system architecture is shown in Fig 3.

The destination information from drivers is necessary for estimating the traffic flows and their distribution across the network. During or before the event, the system polls the destination from divers. The driver destinations are used to build the traffic flows, where the system divides the city into a set of destination areas. All the drivers going to the same destination area are aggregated in the same traffic flow. The number of vehicles in each flow is used to estimate the flow’s initial average traffic rates based on exponential distribution headway intervals. These rates will be used as initial rates by the optimization module and can be updated by the VDC based on the network conditions.
and random interactions between vehicles. To reduce this noise, the system smooths the received link travel time with the link’s previously smoothed value using exponential smoothing (in our case, we used an exponential damping factor $\alpha = 0.2$), as shown in (1).

$$T_{i,j}^t = \alpha T_{i,j}^{t-1} + (1 - \alpha) \hat{T}_{i,j}$$

(1)

3) Estimating the Current Link Loads: The second important parameter to compute the current traffic load $\zeta_{i,j}$ on each link, which is the difference between the link capacity and its current traffic load. The average traffic load on a link can be estimated using the time-mean speed on the link, as shown by the fundamental relationship shown in Fig. 1. Assuming that connected vehicles on the network is a representative sample for vehicles, their time-mean speed can be used as an estimator for the total link time-mean speed. In literature there are models to represent these fundamental relationships such as the Greenshields [29], Pipes [30], Gipps [31], and Van Aerde [32] models.

In our system, we use the Van Aerde model since it combines both the Greenshields and Pipes models. So, the data processing module uses the smoothed $T_{i,j}^t$ and the link length to compute the average link speed $u_{i,j}^t$. Then, it plugs this value in Eq. (2) to compute the average heady distance which is the inverse of the vehicle density.

$$h_{i,j} = c_1 + c_3 u_{i,j}^t + \frac{c_2}{u_f - u_{i,j}^t}$$

(2)

In 2, $c_1, c_2,$ and $c_3$ are computed based on the road network parameters, as:

$$c_1 = \frac{u_f}{K_c u_c^2} (2u_c - u_f)$$

(3)

$$c_2 = \frac{u_f}{K_c u_c^2} (u_f - u_c)^2$$

(4)

$$c_3 = \frac{1}{Q_{i,j}} - \frac{u_f}{K_c u_c^2}$$

(5)

where $u_f$ is the free-flow speed of the link; $u_c$ is its speed-at-capacity; and $K_c$ is the roadway jam density [32]. The headway distance computed in 2 is the reciprocal of the vehicle average density. The traffic load on the link is computed as the multiplication of the link speed by the vehicle density [33], i.e., $\zeta_{i,j} = u_{i,j}/h_{i,j}$.

C. Optimization Module

The optimization module is the core of the proposed system that steers both the VDC and TTSON. Based on the information collected or computed in the previous subsections, the optimization module uses the linear problem described in our previous work [34] to jointly compute allowable vehicle departure rates and optimize their routing.

An Example for Navigation Optimization: We show the idea of optimizing traffic assignment by giving the example in Fig. 4, where a simple network of highway stretch (HW) and an alternative longer travel time arterial road (AR). At point ‘A’, the capacity of the HW is reduced from three to two lanes. Since the HW has a shorter travel time, all vehicles will take it. In the case of high steady-state traffic rate, congestion will take place at point ‘A’ and will spill back, creating a queue on the HW, while leaving the AR underutilized. Using the shortest path routing, this congestion will continue and the travel time of the HW will increase until it becomes longer than that of the AR. At this point, vehicles will start switching to the AR, causing congestion on it. This congestion will switch between the HW and the AR.

In our proposed system, the TTSON will divide the traffic flow between the HW and the AR in such a way that minimizes the network-wide travel time. For instance, if the traffic flow rate is $\mu$ veh/h, then $\beta \mu$ veh/h will take the HW, and $\gamma \mu$ veh/h will take the AR, where $\beta$ and $\gamma$ are in $[0, 1]$, and in this example, $\beta + \gamma = 1$. We call $\beta$ and $\gamma$ the link-flow assignment parameters that must be computed to minimize the network-wide travel time while maintaining the total flow on each road within capacity.

1) Optimization Problem and Constraints: As illustrated by the simple scenario, the optimization module attempts to reduce the network-wide travel time and congestion by utilizing all the available road capacities. To fulfill this objective, a linear optimization model in 6 is formulated. The general idea of the linear program is to divide each flow across a set of alternative routes by computing $q_{i,j}^{k,i}$ that minimizes the network-wide travel time while respecting the road capacities. $q_{i,j}^{k,i}$ is the portion of this traffic flow $k$ that should go through $L_{i,j}$.

$$\min q_{i,j}^{k,i} \sum_{i=1}^{n} \sum_{j=1}^{n} T_{i,j} \sum_{k=1}^{f} q_{i,j}^{k,i}$$

subject to:

$$\sum_{d=1}^{n} q_{i,j}^{k,d} - \sum_{s=1}^{n} q_{i,s}^{k,i} = 0, \text{ if } i \text{ is an intermediate node}$$

$$\sum_{d=1}^{n} q_{i,j}^{k,d} - q_{k} = 0, \text{ if } i \text{ is the source of the } k^{th} \text{ flow}$$

$$q_{k} - \sum_{d=1}^{n} q_{i,j}^{k,d} = 0, \text{ if } i \text{ is the destination of the } k^{th} \text{ flow}$$

$$\zeta_{i,j} + \sum_{k=1}^{f} q_{k}^{i,j} \leq Q_{i,j} \quad \forall \ L_{i,j} \in \mathcal{L},$$

$$q_{k}^{i,j} \geq 0.$$  

(6)

In this model, $i, s, d \in \mathcal{N}$ such that $L_{i,d}$ is a link exiting node $i$, and $L_{s,i}$ is a link entering node $i$. $T_{i,j}$ is the travel time for link $L_{i,j}$, $Q_{i,j}$ is the link capacity, and $\zeta_{i,j}$ is the total traffic load on that link.
The linear program constraints are divided into three sets: route continuity, link capacity, and positive assignment constraints.

The Route Continuity Constraints: This set of constraints is represented by the first three constraints in 6. The objective of these constraints is to ensure that the computed traffic-flow assignment creates a complete and connected route for each flow \( k \) from its origin to its destination. The route continuity condition is achieved by enforcing the individual flow balance at each node, which can be formulated as follows.

- For each intermediate node \( i \), and for each flow \( k \), the summation of the traffic of the \( k^{th} \) flow entering this \( i^{th} \) node must be equal to the summation that is exiting this node.
- For each source (or destination) node, we add (or subtract) the total flow rate \( q_k \). For instance, for the source node \( i \) that generates the \( k^{th} \) flow whose rate is \( q_k \), we assume there is a fictitious source sending \( q_k \) to it, then node \( i \) sends these vehicles to other nodes \( d \in N \). And vice versa for the destination nodes.

The Link Capacity Constraints: Constraints in this set are used to control the load on the roads by limiting the total traffic traversing a link \( L_{i,j} \in L \) to its capacity \( Q_{i,j} \). The total traffic rate on the link is calculated as the summation of its current load \( \zeta_{i,j} \) and all the traffic assigned to this link \( \sum_{k=1}^{f} q_{i,j}^{k} \).

The Positive Assignment Constraints: The last constraint is to allow only positive values for \( q_{i,j}^{k} \), to make it consistent with the directed links.

Link costs \( T_{i,j} \) and traffic load \( \zeta_{i,j} \) are continuously updated using the methodology described earlier. The system periodically solves this problem to cope with the dynamic network condition. This period is called Re-Optimization Interval (ROI).

D. Navigation Module

When a vehicle requests a route (or a route update), the communication module forwards this request to the TTSON. To build a route for a vehicle, TTSON, finds the parameters needed to build the route, namely; the vehicle’s destination node \( d \), the vehicle’s flow number \( k \) (inferred from its destination), its current location from which it can find the route starting node \( s \), and all the traffic assigned to this link \( \sum_{k=1}^{f} q_{i,j}^{k} \).

The Positive Assignment Constraints: The last constraint is to allow only positive values for \( q_{i,j}^{k} \), to make it consistent with the directed links. The link costs \( T_{i,j} \) and traffic load \( \zeta_{i,j} \) are continuously updated using the methodology described earlier. The system periodically solves this problem to cope with the dynamic network condition. This period is called Re-Optimization Interval (ROI).

Algorithm 1

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>procedure Build a vehicle route ( k, q_{i,j}^{k}, s, d )</td>
</tr>
<tr>
<td>2.</td>
<td>( R \leftarrow \phi )</td>
</tr>
<tr>
<td>3.</td>
<td>( i \leftarrow s )</td>
</tr>
<tr>
<td>4.</td>
<td>while ( i \neq d ) do</td>
</tr>
<tr>
<td>5.</td>
<td>( R \leftarrow (R, i) )</td>
</tr>
<tr>
<td>6.</td>
<td>( \hat{L} \leftarrow L_{i,j}: L_{i,j} \in L )</td>
</tr>
<tr>
<td>7.</td>
<td>for each ( L_{i,j} \in \hat{L} ) do</td>
</tr>
<tr>
<td>8.</td>
<td>compute ( p_j )</td>
</tr>
<tr>
<td>9.</td>
<td>( r \leftarrow \text{randomnumber} )</td>
</tr>
<tr>
<td>10.</td>
<td>for ( j = 1 ) to size of ( \hat{L} ) do</td>
</tr>
<tr>
<td>11.</td>
<td>if ( \sum_{j=1}^{f} p_j \leq r ) then</td>
</tr>
<tr>
<td>12.</td>
<td>( i \leftarrow j )</td>
</tr>
<tr>
<td>13.</td>
<td>break</td>
</tr>
<tr>
<td>14.</td>
<td>return ( R ) ( \triangle ) Return the computed route.</td>
</tr>
</tbody>
</table>

E. Departure Control Module

The VDC module is responsible for regulating the traffic entering the network to avoid network congestion. Moreover, in the case of high traffic demand (i.e., when the traffic demand rates exceed the available network maximum flow rate), the VDC computes the maximum allowable traffic rates \( \hat{q}_{k} \leq q_{k} \), and, accordingly, adjust the vehicle departure times.

1) Computing the Maximum Allowable Traffic Rates: There are existing algorithms to compute the multi-origin multi-destination maximum flow, such as [35], [36]. But these solutions have two main drawbacks. First, the solution they achieve may result in a complete blockage of some traffic flows (i.e., setting \( \hat{q}_{k} = 0 \) for some flows), which violates the fairness among the traffic demands. Additionally, this flow blockage will result in differing vehicles in these flows for a long time, consequently, increasing the network clearance time. Secondly, in multi-origin multi-destination maximum flow algorithms, the computed traffic rates do not necessarily minimize the travel time since the maximum flow algorithms do not consider other metrics other than road capacities.

To maintain fairness and minimize travel time, the VDC determines the network maximum flow rates by reducing all the traffic demand rates by the same ratio, and then find if there is a feasible solution for the linear problem in 6 using these rates. It relies on optimization failure as an indication of high traffic demand. To find the maximum allowable departure rates, the VDC applies the algorithm shown in the flowchart in Fig. 5. The algorithm is similar to the binary search in a sorted list. \( R_{\text{Max}} \) and \( R_{\text{Min}} \) are the maximum and minimum boundaries of the search space that are updated each iteration. The \( R_{\text{Max}} \) is initialized to the initial rate computed based on the drivers leaving times. \( R_{\text{Cur}} \) is the current traffic rates, it is initialized...
After computing the current time. This equation shrinks or extends the vehicle inter-departure intervals by the ratio $\frac{q_k}{\bar{q}_k}$.

The departure time is sent to drivers enough time before his/her departure. The VDC uses a token-based model to control the departure of vehicles. So, if the departure time for vehicle $v$ is computed to be $t_v$, the VDC sends it a token at time $t_v - T_p$, where $T_p$ is the preparation time, i.e., the time needed for the driver to go to the car, to leave the parking lot and go to the gate at designated exit points (i.e., G1, G2, G5 in Fig. 3). Once the VDC has sent a token to a driver, the VDC cannot change their departure time.

V. SIMULATION AND RESULTS

To accurately test the proposed model, we developed it within the INTEGRATION software [37], which is an agent-based microscopic traffic simulation and assignment framework. It is characterized by its accuracy in computing travel time and capturing network congestion and its impact on travel time. INTEGRATION can accurately compute travel time because it replicates vehicle longitudinal motion based on the Van Aerde model [37], which produces the best regression for real datasets compared to other models [38]. In INTEGRATION, vehicle speed and acceleration are constrained by the vehicle dynamics [37]. Moreover, its microscopic nature enables it to accurately model vehicle queues and shock-wave at traffic lights [39], which significantly affects the link travel time. It also accounts for traffic lights and their impact on travel time.

We use the CPLEX Optimizer [40] to compute the traffic assignment. To study the efficiency of the proposed framework, we compare it to the Sub-population Feedback Dynamic Time-Based Traffic Incremental Assignment (SFDTIA) [37], which uses the shortest path routing based on the dynamic link travel times since this would best reflect the state-of-practice routing. SFDTIA also tries to overcome the shortest path routing problem and utilize the network resources by dividing the traffic into five sub-populations; each sub-population is routed differently at the same time. This way, it can utilize alternative routes [37].

We use the 2022 FIFA World Cup event, which will be held in Doha, Qatar as a case study. After each match, the number of vehicles that need to leave the stadium will be huge, resulting in a vehicle crowd surrounding the stadium area. The following subsection describes the network and traffic demand setting.

### A. Network and Traffic Setup

To create a real-world scenario of the 2022 World Cup case study, the Doha road network, shown in Fig. 2, is developed and used for the simulation.

To build the simulation network and to generate the associated parameters (such as the road speed, number of lanes on the road,
and traffic lights), we used data from different sources, namely:
1) a road network Geographic Information System (GIS) Shape-
file; 2) intersection data from OpenStreetMap (OSM); and 3)
Google maps and ArcGIS. The four key traffic stream param-
ters: free-flow speed, speed-at-capacity, saturation flow rate, and
jam density were also derived from different sources. Specifi-
cally, the free-flow speed was derived from the roadway speed
limits. The speed-at-capacity was set at 80% the free-flow speed
as has been demonstrated in the literature [41], [42], the satura-
tion flow rate was derived using the Highway Capacity Manual
(HCM) based on the roadway free-flow speed. Finally, the jam
density was computed using a typical vehicle length. The Doha
city shapefile was used to generate the network nodes and links.
OpenStreetMap data were used to extract intersection traffic
control information including the traffic control methods (stop
sign, yield sign, or traffic signals). The number of phases for each
traffic signal and traffic signal timing data was obtained based
on field observation and was augmented with real-time traffic
signal optimization. Google maps and ArcGIS were utilized
for validating road attributes, including the number of lanes,
one-way streets, and speed limits for each road segment. The
resulting simulation network has 169 nodes, 301 road segments,
and 11 traffic signals.

The red area in Fig. 2 represents a stadium from which
vehicles belong to the controlled traffic (CT) will depart to
different network points (as shown by the yellow arrows). This
traffic is distributed over ten network exit points.

In regards to background traffic (BT), which represents the
regular traffic traversing the network, it was calibrated based
on car counts data collected from the OpenStreetMaps (OSM)
website. The BT rate in the first row \( S_1 \) in Table II is 10% of
the calibrated traffic. In this scenario, \( S_1 \), we assume that there are
2800 cars as the CT. The rates in \( S_1 \) are multiplied by scaling
factors 2 through 5 to compute the higher traffic rates in \( S_2 \)
through \( S_5 \). We assume that these vehicles should leave within
one hour with uniform inter-departure intervals. This way, these
car counts can be converted into traffic rates.

### B. Simulation Results

Each of the five traffic levels cases was run using both
three methods; the Base case (SFTDIA), the TTSON, and TT-
SON+VDC. For statistically significant results, each scenario
was run 16 times with different seeds and the average parameters
computed.

#### 1) Network Clearance Time

Fig. 6 compares the network clearance time for the TTSON, TTSON with VDC, and the
SFTDIA for the five traffic levels.

It shows that the TTSON can clear the network faster in high
traffic demand levels, which are typical cases of its use. In the
case of low traffic rates, the differences in the clearance time are
not significant. It also illustrates that the higher the traffic rate
the better the saving it achieves.
By applying the VDC with the TTSON, Fig. 6 also demonstrates an improvement in the network clearance time in most cases, while in some cases, the VDC slightly increases the network clearance time. The reason is that the final impact of the VDC has two opposing components. The first is that by reducing the traffic load, it reduces the congestion and decrease vehicle travel times, which allows the network to clear early. On the other hand, by decreasing the traffic load, vehicles start their trips late, which may increase the clearance time.

The figure also shows that, at the low traffic levels (S1, S2, and S3), the VDC does not have any impact. The reason is that, at these low rates, the network capacity is sufficient for the traffic. Thus it does not need to decrease the traffic rates.

An interesting observation in Fig 6 is that, in some cases, longer ROIs result in shorter network clearance time. For instance, in the case of S3, increasing the ROI from 60 sec to 120 sec results in 3% reduction in the network clearance time. The reason is that while lowering the ROI helps the CT to arrive early, it may increase the trip time for the BT (as we will show later in Section V-B4 and Fig. 10). Consequently, in some cases, the delay experienced by the BT may increase the total network clearance time.

2) Average Trip Time: The trip time for each individual vehicle is the time it takes from its scheduled departure time to the moment it reaches its destination. Fig. 7 compares the system-wide average travel time, which is computed as the summation of all vehicles’ travel times divided by the number of vehicles.

Fig. 7 demonstrates that at low traffic levels, the proposed system increases the average trip time by around 10%. This increase is attributed to the sensitivity of the system to congestion on road segments, which steers the TTSON to route vehicles through longer routes. Consequently increases the travel time for those vehicles.

However, as the traffic rate increases, the proposed system shows significant improvements in the system-wide average trip time that exceed 50% in extreme cases.

3) Time-Space Diagram: Fig. 8 shows the time-space diagram for 0.5% of the cars for the two high traffic demand levels. Each line in this figure represents a vehicle trajectory (the distance it travelled versus time). It shows that for the base case, many vehicles departed very late. Despite the fact that these vehicles attempted to depart in the first hour, we can see some vehicles departed after 160 minutes, implying that these trips were delayed for more than an hour. The reason is that SFDTIA uses a limited number of alternative shortest routes, thus when a driver tries to start their trip and finds its entry road full, it defers its departure. While in our proposed system, a car can be assigned longer routes in which the entry links are almost underutilized. Thus, not requiring the deference of vehicle departure times. This is clear in the cases of TTSON and TTSON with VDC, where the last car departed within 80 minutes of the simulation time. The second point in this figure is that the car speeds in the case of SFDTIA are higher, which is reflected by the steep trajectory slopes, which is attributed to the low vehicle density in the network.

A time-space diagram also shows that in the case of the TTSON without the VDC, there is severe congestion in the network, which can be depicted from the horizontal blue lines. Such horizontal lines demonstrate that a vehicle is stuck in congestion; consequently, it does not move. Because of this congestion, the TTSON system detours some cars through much longer routes (shown by the long blue lines).
By employing the VDC with the TTSON, the system is able to reduce the congestion. Therefore, reducing or eliminating the need for longer detours. Consequently, the VDC can reduce the average trip total travel time as shown in Fig. 7.

4) No Free Lunch, CT vs. BT: We also analyzed the traffic class specific performance for the two classes; CT and BT. Fig. 9 and Fig. 10 show the average trip time for CT and BT, respectively. Fig. 9 illustrates the significant impact of the proposed system on the CT average trip time. It shows that the TTSON significantly reduced the CT average trip time. The figure also demonstrates that using the VDC system, contributed to more reduction in the average CT trip time at the high traffic demand levels S4 and S5. It also shows that the impact of the VDC increases with the traffic demand, which is aligned with its purpose of regulating traffic in the congested cases. These improvements are achieved at the cost of the BT, as shown in Fig. 10, where it shows a significant increase in the BT average trip time.

An interesting observation in Fig. 10 is that the VDC results in longer BT average trip time, which contrasts with regulating the vehicle departures. This observation inspired us to investigate this further. By comparing the vehicle density in the network in these different cases, it turned out that regulating the traffic at the gates reduces the congestion at the exit links, consequently allows vehicles to enter the network faster. Thus, resulting in higher vehicle density in the network as shown in Fig. 11 that adversely affects the BT trip time which was doubled in some cases.

This also shows the importance of the TTSON in efficiently controlling the traffic assignment, where the CT can be routed in a better way that decreases its travel time despite the higher vehicle density compared to the SFDTIA.

VI. CONCLUSION
In this paper, we present a vehicle crowd management system, entitled IoT-VCM, that optimizes the performance of the transportation system in the case of high vehicular demand levels.

To optimize the road network performance and clear the network faster, two conflicting objectives exist, namely; early
vehicle departures (which results in higher congestion levels) and short travel times (which requires low traffic congestion). Our proposed system addresses these contradictory objectives by allowing vehicles to depart as early as possible while constraining the traffic load to the maximum network flow rate. Then, to minimize the overall network-wide total travel time, vehicles are routed using a dynamic system-optimimum traffic assignment.

The microscopic simulation of the proposed system shows its ability to clear the network in a shorter time while maintaining the travel time as low as possible. The analysis of the vehicle trajectories shows lower congestion when employing the vehicle departure control. Moreover, in some cases, regulating the vehicle traffic entering the network using the vehicle departure control allows vehicles to enter the network faster and increase the vehicle density in the network while still having lower trip times.

REFERENCES

Ahmed Elbery (Member, IEEE) received the B.Sc. degree (with Hons.) in electrical and computer engineering from the Department of Electrical Engineering, Benah University, Egypt, and the M.Sc. degree in electronic engineering from the Department of Electronic Engineering, Menoufia University, Egypt. He received the Ph.D. degree in computer science from Virginia Tech, USA. He is a Postdoctoral Research Fellow with the School of Computing, Queen’s University. His research interests include communication and data networking, wireless and vehicular communication, ITS, simulation and modeling of large-scale vehicular networks in smart cities, eco-routing navigation systems. He was a Senior Network Engineer and has numerous network and communication certificates from different vendors including Cisco, Juniper, and Huawei-3Com. He also has network instructor certificates from Juniper, JNCIA, and JNCIS Junos Enterprise Routing and Switching track. He is a Cisco Academy instructor for CCNA, CCNA Wireless, and CCNA Security.

Hossam S. Hassanein (Fellow, IEEE) is a leading authority in the areas of broadband, wireless and mobile networks architecture, protocols, control and performance evaluation. His record spans more than 500 publications in journals, conferences and book chapters, in addition to numerous keynote speeches and plenary talks in flagship venues. Dr. Hassanein was the recipient of the several recognition and Best Papers Awards at top international conferences. He is also the Founder and Director of the Telecommunications Research Lab (TRL), Queen’s University School of Computing, with extensive international academic and industrial collaborations. He is a Former Chair of the IEEE Communication Society Technical Committee on Ad hoc and Sensor Networks (TC AHSN). Dr. Hassanein is an IEEE Communications Society Distinguished Speaker.

Hesham A. Rakha (Fellow, IEEE) received the Ph.D degree from Queen’s University, Kingston, Ontario, in 1993. He is currently the Samuel Reynolds Pritchard Professor of Engineering with the Department of Civil and Environmental Engineering and the Department of Electrical and Computer Engineering (Courtesy), Virginia Tech, and the Director of the Center for Sustainable Mobility with the Virginia Tech Transportation Institute. His research interests include large-scale transportation system optimization, modeling, and assessment. He works on optimizing transportation system operations, including vehicle routing, developing various network and traffic signal control algorithms, developing freeway control strategies (speed harmonization and ramp metering), and optimizing vehicle motion (lateral and longitudinal control of connected automated vehicles (CAVs)) to enhance their efficiency and reduce their energy consumption while ensuring their safety. He has authored or coauthored of six conference Best Paper Awards, namely: 19th ITS World Congress (2012), 20th ITS World Congress (2013), VEHITS (2016), VEHITS (2018), and TRB (2020), received the most cited paper award from the International Journal of Transportation Science and Technology (IJTST) in 2018 and received 1st place in the IEEE ITSC 2020 UAS4T Competition. In addition, Dr. Rakha was the recipient of the Virginia Tech’s Dean’s Award for Outstanding New Professor (2002), the College of Engineering Faculty Fellow Award (2004–2006), and the Dean’s Award for Excellence in Research (2007). He is an Editor for Sensors (the Intelligent Sensors Section), an Academic Editor for the Journal of Advanced Transportation, an Associate Editor for the IEEE TRANSACTIONS OF INTELLIGENT TRANSPORTATION SYSTEMS and the Journal of Intelligent Transportation Systems: Technology, Planning and Operations. Furthermore, he is on the Editorial Board of the Transportation Letters: The International Journal of Transportation Research and the International Journal of Transportation Science and Technology.

Nizar Zorba (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from JUST University, Jordan, in 2002, and the Ph.D. degree in signal processing for communications from UPC Barcelona, Spain, in 2007. He is a Professor with the Electrical Engineering Department, Qatar University, Doha, Qatar. He has authored five international patents and coauthored more than 120 papers in peer-reviewed journals and international conferences. He is Associate/Guest Editor for the IEEE COMMUNICATIONS LETTERS, IEEE ACCESS, IEEE Communications Magazine and IEEE NETWORK. He is Symposium Chair at IEEE Globecom 2021, IEEE VTC 2020 and IEEE ICC 2019; Workshop Chair at ICC 2021, IWCMC 2020 and CSIM-PS 2019. He is currently the Vice-Chair of the IEEE ComSoc Communication Systems Integration and Modeling Technical Committee (TC CSIM).