Evolution of Road Traffic Congestion Control: A Survey from Perspective of Sensing, Communication, and Computation

Wenwei Yue¹², Changle Li¹²*, Guoqiang Mao¹², Nan Cheng¹², Di Zhou¹

¹ State Key Laboratory of Integrated Services Networks, Xidian University, Xi’an 710126, Shaanxi, China
² Research Institute of Smart Transportation, Xidian University, Xi’an 710126, Shaanxi, China
* The corresponding author, email: clli@mail.xidian.edu.cn

Abstract: Road traffic congestion can inevitably degrade road infrastructure and decrease travel efficiency in urban traffic networks, which can be relieved by employing appropriate congestion control. According to different developmental driving forces, in this paper, the evolution of road traffic congestion control is divided into two stages. The ever-growing number of advanced sensing techniques can be seen as the key driving force of the first stage, called the sensing stage, in which congestion control strategies experienced rapid growth owing to the accessibility of traffic data. At the second stage, i.e., the communication stage, communication and computation capability can be regarded as the identifying symbols for this stage, where the ability of collecting finer-grained insight into transportation and mobility reality improves dramatically with advances in vehicular networks, Big Data, and artificial intelligence. Specifically, as the prerequisite for congestion control, in this paper, existing congestion detection techniques are first elaborated and classified. Then, a comprehensive survey of the recent advances for current congestion control strategies with a focus on traffic signal control, vehicle route guidance, and their combined techniques is provided. In this regard, the evolution of these strategies with continuous development of sensing, communication, and computation capability are also introduced. Finally, the paper concludes with several research challenges and trends to fully promote the integration of advanced techniques for traffic congestion mitigation in transportation systems.

Keywords: road traffic congestion control; congestion detection; traffic signal control; vehicle route guidance; sensing techniques; communication and computation capability

I. INTRODUCTION

1.1 Road Traffic Congestion and Need for Traffic Control

In recent decades, road traffic congestion has experienced unprecedented growth in almost all metropolitan areas because of the dramatic development of traffic demands for all transportation modes. The Texas Transportation Institute reported that traffic congestion caused a waste of time and fuel consumption valued at over $150 billion during 2014 for the 470 metropolitan areas in the United States, which was projected to expand to almost $190 billion in 2020 [1, 2]. Additionally, on the basis of the Urban Mobility Scorecard in 2015, road traffic congestion caused over three billion gallons of extra fuel dissipation during 2014, which led to more CO₂ emissions and serious greenhouse effects [1]. More importantly, the negative impacts of traffic congestion are becoming increasingly severe, as shown in Figure 1, and the travel delays of each auto commuter are rapidly increasing in cities of all sizes. That is to say, congestion is not just a problem of big cities; an ever-growing number of urban areas around the world are also experiencing increasingly severe traffic conditions.
To relieve these negative impacts, congestion mitigation strategies are extensively adopted by worldwide transportation authorities. According to an urban transportation report for 498 U.S. urban areas, congestion mitigation strategies are gauged to reduce more than 370 million hours of vehicle travel time and save $8.5 billion in 2011 by using a combination of real-time traffic information, control technologies, operating practices, and construction programs [3]. If these strategies can be adopted on most roadways, the benefits are expected to decrease more than 800 million hours of travel time valued at $20 billion [3]. These significant benefits have encouraged both industry and academia to make continuous research and development efforts for congestion control and mitigation.

1.2 Evolution of Congestion Control Strategies

The capability of detecting road congestion is the premise of traffic control and management. However, estimating the traffic situation of an urban road network is a difficult problem due to its intrinsic characteristics: 1) traffic conditions are inherently stochastic; and 2) the type, accessibility, coverage, and quality of real-time data vary significantly [4]. As the expansion of data collection techniques in intelligent transportation systems (ITSs), a large amount of advanced sensing, communication, and intelligence technologies are investigated and employed in traffic congestion detection, such as video cameras [5], vehicular networks [6], and social media [7], which provide many promising approaches for understanding congestion in urban networks with high accuracy and low latency.

Among congestion mitigation strategies, traffic signal control and route choice are the most extensively applied and perhaps the most effective ways of controlling and relieving congestion. After the appearance of elementary traffic signal control and vehicle guidance strategies in the last century, traffic signal control and vehicle route guidance have undergone continuous improvements. In general, the evolution of these traffic congestion control methods can be classified into two stages: sensing and communication, as illustrated in Figure 2. The key driving force for the first stage is the ever-growing advanced sensing techniques. Specifically, the earlier traffic light controllers adopted fixed-time control methods by using pre-determined cycles and split time plans, such as the traffic network study tool (TRANSYT) [8]. After that, a number of traffic-responsive signal control strategies were achieved gradually under the trend of advanced sensing techniques, such as the Split, Cycle and Offset Optimization Technique (SCOOT) [9] and Sydney Cooperative Adaptive Traffic System (SCATS) [10], which can better respond to the prevailing traffic conditions and relieve traffic congestion effectively. For vehicle route guidance, the re-routing methods also evolved from deterministic to adaptive strategies utilizing the real-time traffic information provided by advanced sensing techniques, which greatly improved the traffic efficiency considering traffic uncertainty.

Since the 2000s, traffic congestion mitigation strategies have undergone vigorous development, which has dramatically enhanced the ability for grasping more nuanced insight into transportation systems [11] with the development of vehicular networks, Big Data, and artificial intelligence. All promote traffic control research into the next stage, in which the intelligence and communication capability can be regarded as the identifying symbols [12, 13]. First, during the last 25 years, data have increased tremendously in various fields, especially within transportation systems. The volume and variety (source, type, and format) of transportation data raise many opportunities for current ITS administrators to conduct effective traffic monitoring, decision-making, and data management. Meanwhile,
artificial intelligence provides advanced algorithms in terms of 1) collecting and analyzing the massive real-time traffic data rapidly, and 2) conducting pattern recognition, prediction, and classification for different applications in ITSs. In particular, vehicular networks act as a “porter” to maintain the computation process, and transmit traffic data to places with computing capacity and advanced algorithms. In summary, this communication stage synthesizes functions for the acquisition of massive, heterogeneous, and real-time data and the applications of collected traffic data by employing the advanced information analytical algorithms, which enables transport issues to be conducted from a pioneering perspective and lays the foundation for more intelligent and efficient transportation systems in the future [14].

1.3 Contributions and Organization

This paper is aimed at elaborating the evolution of key congestion control technologies in both sensing and communication stages. Unlike existing surveys about road congestion control that concentrated on the applications of advanced techniques for traffic signal control [13, 15, 16] or vehicle route choice [17], the paper provides a comprehensive review for urban congestion control strategies from the perspective of sensing, communication and computation, which can not only present a reference for researchers from different fields to understand the key technologies and development process of this research area, but also share the novel ideas about state-of-the-art techniques, challenges and future research with transportation experts to raise more new issues and challenges in this topic. In particular, the contributions of this paper are the following. First, congestion detection methods are introduced and classified according to different data collection and sensing techniques. The pros and cons of these techniques for congestion detection are discussed. Next, as extensively applied congestion mitigation strategies, traffic signal control, vehicle route guidance, and their combined techniques are introduced. Their development processes based on evolving advanced sensing, communication, and computation capability are also presented. Finally, detailed discussions of the trends and future research directions of traffic congestion control are provided.

The rest of this paper is organized as follows. In Section II, existing traffic congestion detection methods utilizing different sensing techniques are intro-
duced. In Sections III, IV, and V, reviews of traffic signal control, route guidance, and their combined techniques are presented, respectively. Discussions of challenges and future research directions are highlighted in Section VI. Finally, Section VII concludes the paper.

II. TRAFFIC CONGESTION DETECTION

As a pre-requisite to relieving traffic congestion, traffic congestion detection is becoming increasingly important. Based on the detected congestion-related information, many congestion mitigation strategies, such as traffic signal control and vehicle rerouting, can be utilized to relieve congestion. To collect traffic information and monitor traffic congestion, stationary roadside sensors are widely used. The inductive loop detector is among the most applied technique for congestion detection to monitor real-time travel speed and road density [18–20]. In the last decade, vision-based vehicle detection technologies have become an extremely active research area; they use image processing techniques to analyze and calculate road density for traffic congestion detection [21, 22]. Many other stationary roadside sensors, such as radio-frequency identification (RFID) readers [23] and magnetic sensors [24] are also used in traffic information acquisition and have demonstrated their effectiveness in traffic congestion detection. Moreover, as another widely used road information collection technique, probe-vehicle-based techniques also play vital roles in traffic congestion detection. Probe-based techniques can be classified into two categories: global-positioning-system- (GPS-) based [25, 26] and cellular-signal-based techniques [27]. The widespread use of GPS devices in vehicles provides a feasible solution for urban congestion detection, which aggregates real-time GPS traces from probe vehicles to analyze and estimate traffic conditions. With the prevalence of mobile phones, congestion detection using cellular-signal-based techniques have also drawn much attention, and utilizes embedded sensors such as accelerometers (road traffic condition recognition) and cellular signals (map matching) to detect road traffic congestion in urban areas. In recent years, researchers have investigated the use of vehicular networks to discover traffic congestion information [28, 29]; such networks collect traffic information from surrounding vehicles periodically via vehicular communications and monitor real-time traffic conditions on road segments. Furthermore, due to the rapid development of mobile devices, many works have been developed to detect congestion based on social media platforms (e.g., Twitter and WeChat) by using syntactic analysis [30, 31]. The comparisons of these congestion detection techniques are presented in Table 1.

2.1 Stationary Roadside Sensors

Considering the huge installation and maintenance costs for stationary roadside monitors, these sensors cannot be installed on all roads in urban traffic networks. Therefore, for the road segments with installed sensors, the traffic information (e.g., travel speed, road occupancy, and traffic flow etc.) can be measured directly. For road segments without installed roadside sensors, however, the traffic information can be estimated indirectly using traffic information of other road segments with sensors and traffic correlation between road segments [39, 40].

2.1.1 Direct Detection

Stationary roadside sensors, such as inductive loop detectors and image processing and magnetic sensors are among the most widely accepted methods for traffic measurements. Based on these roadside sensors, many approaches have been proposed to acquire traffic information and estimate congestion on road segments, which provides a significant premise for real-time traffic control and congestion mitigation in traffic networks.

**Inductive loop detectors:** As one of the most commonly used roadside detectors [41], inductive detection loops are deployed across the roadway and occupy a major portion of the lanes where traffic detection is required. There are two main categories of loop detectors: single- and double loop detectors [42]. Single-loop detectors are more suitable for vehicle counting, while double-loop detectors are fit to capture vehicle travel speeds. Cherrett et al. [43] utilized single-inductive-loop detectors to calculate the average loop occupancy time per vehicle and the average time gap between vehicles, which were used to detect congestion on the detected road segments. In [44], Lao et al. took the ratio of traffic demand (traffic flow rate) and traffic supply (road capacity) into consideration
### Table 1. Comparison of traffic congestion detection techniques.

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stationary Road-side Sensors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inductive loop detector</td>
<td>• Mature techniques</td>
<td>• High installation and maintenance costs</td>
</tr>
<tr>
<td></td>
<td>• Easy to implement</td>
<td>• Limited coverage</td>
</tr>
<tr>
<td></td>
<td>• Low production cost</td>
<td>• Damage on road surface [32]</td>
</tr>
<tr>
<td>Video camera</td>
<td>• Collect more traffic parameters</td>
<td>• Sensitive to weather condition and scene changes</td>
</tr>
<tr>
<td></td>
<td>• Captured pictures can be used repeatedly</td>
<td>• Limited coverage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Vast storage space and powerful computing capability [33]</td>
</tr>
<tr>
<td>Magnetic sensor</td>
<td>• Robust in extreme weather conditions</td>
<td>• Cannot detect static vehicles</td>
</tr>
<tr>
<td></td>
<td>• Energy-efficient with wireless transmission and easy to install or maintain</td>
<td>• Limited coverage</td>
</tr>
<tr>
<td></td>
<td>• Effective when working with other sensors or as a node in a sensor network</td>
<td></td>
</tr>
<tr>
<td><strong>Probe-based Techniques</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPS based technique</td>
<td>• Less installation and maintenance costs</td>
<td>• Limited representation</td>
</tr>
<tr>
<td></td>
<td>• More accurate and timely in vehicle localization</td>
<td>• Power hungry</td>
</tr>
<tr>
<td></td>
<td>• High coverage</td>
<td>• GPS localization performance is not stable in urban areas [35]</td>
</tr>
<tr>
<td>Cellular signal based technique</td>
<td>• Less installation and maintenance costs</td>
<td>• Inaccuracy in estimating vehicle positions and motions</td>
</tr>
<tr>
<td></td>
<td>• High coverage</td>
<td>• Difficulty in computing travel speed [26]</td>
</tr>
<tr>
<td></td>
<td>• High reliability in vehicle localization</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• High sampling representation</td>
<td></td>
</tr>
<tr>
<td>Radar/Lidar based technique</td>
<td>• Direct distance measurement</td>
<td>• Difficulty in detecting highly reflective objects</td>
</tr>
<tr>
<td></td>
<td>• Effective when detecting obstacles in road environment</td>
<td>• Sensitive to radars and lidars measurement latency</td>
</tr>
<tr>
<td></td>
<td>• Good performance in poor visibility environments</td>
<td>• Poor very near (&lt; 2m) measurement [36]</td>
</tr>
<tr>
<td><strong>Vehicular Networks</strong></td>
<td>• Less installation and maintenance costs</td>
<td>• V2X message transmission can be restricted by high buildings close to the streets</td>
</tr>
<tr>
<td></td>
<td>• High coverage</td>
<td>• Sensitive to V2X communication latency</td>
</tr>
<tr>
<td></td>
<td>• Easy to deploy and extend</td>
<td>• Communication overhead and packet collisions [37]</td>
</tr>
<tr>
<td></td>
<td>• Low costs and delays in traffic information detection</td>
<td></td>
</tr>
<tr>
<td><strong>Social Networks</strong></td>
<td>• Widely used social networks and media platform</td>
<td>• Status update messages are unstructured and irregular with informal words and misspellings [38]</td>
</tr>
<tr>
<td></td>
<td>• Low costs and delays</td>
<td>• Malicious attacks and trust issue</td>
</tr>
<tr>
<td></td>
<td>• More congestion related information</td>
<td></td>
</tr>
</tbody>
</table>

to detect congestion on urban road segments based on single-loop-detector data, which can be used for congestion quantification and traffic control. In [45], Ali et al. presented a novel inductive loop sensor that can not only sense and classify the vehicle type (e.g., bicycle, car, or bus), but provide precise vehicle counting; they have been used in field tests and shown their effectiveness in detecting vehicles and congestion. Ki
et al. [46] utilized an error-filtering method based on a double-loop detector to enhance the precision for speed metering and congestion detection on road segments. Li et al. [2] identified the congestion bottlenecks based on inductive loop detector data in Taipei urban road networks. Furthermore, in [47], by differentiating congestion effects among road segments and intersections, a congestion propagation based method was developed to attribute the root cause of congestion to road capacity and signal control respectively. Experiments using the loop detector data in Taipei were carried out to identify congestion root causes of the road network, which provides specific guidance for road administrators to mitigate traffic congestion.

**Video camera**: Video camera techniques in vehicle detection have drawn significant attention recently and many efforts have been fulfilled in theory and application [48–51]. Utilizing video cameras, a number of image processing techniques have been proposed to identify and classify vehicles on urban roads, and have provided rich traffic information for congestion detection in urban areas. The main idea of congestion detection based on video cameras is to distinguish foreground objects (moving vehicles on road segments) from stationary backgrounds (e.g., roads, trees, and street-parked vehicles) and calculate the average travel speed and road occupancy for congestion detection. Hsiao et al. [49] proposed a streaming-based method for road congestion detection. Vehicles in each image are detected using the background-extraction method. The encoded vehicles and road segments utilize different bit rates and decode two bit streams for vehicle counting and traffic congestion detection. Friedman and Russell [50] developed a probabilistic technique to separate background and vehicles, wherein an expectation-maximization- (EM-) algorithm-based method was proposed to differentiate moving vehicles and backgrounds. Zhang, Luo, and Zhou [51] presented a background-updating algorithm for road condition monitoring using road videos. They created the corresponding matching relationship of the same vehicles between continuous image frames and predicted the positions of target vehicles on the next frame using Kalman filtering. When the gap between the predicted and actual positions of a vehicle exceeded a prescribed value, this vehicle was considered a “congestion vehicle”, and if the quantity of “congestion vehicles” exceeded a certain threshold, the road could be seen as being congested.

**Magnetic sensors**: Magnetic sensors can identify the movement of vehicles on road segments by measuring magnetic field variation. This fact has been utilized to detect congestion and acquire traffic information [24]. In [52], Cheung et al. proposed a vehicle monitoring technique using roadside magnetic sensors that can achieve accurate vehicle detection and counting. Lan et al. [53] detected and classified traveling vehicles by processing magnetic signals for road segments with deployed magnetic sensors, in which the features of magnetic signals were extracted to estimate the traveling direction and type of vehicles. Researchers in [54] constructed a wireless system to detect vehicle arrivals, departures, and speed using magnetic sensors and accelerometers that can be used in traffic flow and road occupancy estimation for congestion detection. In [32], Wang et al. studied vehicle detection using magnetic sensors. Based on the magnetic signals, wavelet methods were used to remove the negative effects of noise and increase the precision for vehicle surveillance and congestion detection.

### 2.1.2 Indirect Detection

Because of the high costs of roadside data collection facilities, stationary roadside detectors are often deployed on limited numbers of road segments, and are therefore insufficient to meet requirements for traffic data collection in metropolitan areas around the World [39, 40]. Accordingly, for road segments without traffic detectors, traffic congestion must be estimated indirectly utilizing the measurement data on other road segments and their traffic correlations. Most existing works on traffic data imputation were conducted assuming partially missing data, which considers the unavailable data at certain times on road segments because of sensor failure and noise interference [55, 56]. There are also few works focusing on the problem of traffic data imputation considering complete unavailability of traffic data for road segments without sensor deployment. In [40], according to historically available traffic data on road segments with installed sensors, Abadi et al. first generated traffic flows for all road segments in a network using an urban mobility simulator. Then, they optimized and adjusted the route plans for all origin-to-destination pairs and exploited a route plan that minimized the gaps between available
traffic data and the imputed flows on all road segments. Finally, according to the optimized route plans, the traffic flows of road segments without equipped traffic sensors can be estimated. Simulation results showed the effectiveness of the presented technique in missing data imputation for road segments without traffic detector deployment, which also provided a meaningful approach to detect congestion for these road segments.

2.2 Probe-based Techniques

Owing to the extensive coverage and low costs of mobile phones compared to traditional stationary roadside sensors [57], probe-based technologies for traffic surveillance have been seen as one of the most popular topics in mobile sensing and provide an important technique for real-time traffic information collection and traffic condition estimation [26]. According to the different localization techniques used, probe-based road detection methods can be generally categorized as GPS-, cellular-signal-, and radar/lidar-based techniques. GPS-based techniques provide efficient and effective methods for collecting traffic data and congestion detection in urban areas by locating the positions for vehicles and acquiring the real-time travel speeds. Cellular-signal-based techniques utilize cell-tower signal information to identify vehicle locations via tower hand-offs, triangulation, or their combination, and provide more energy-efficient and alternative methods of monitoring traffic information and detecting urban congestion. Radar/lidar-based techniques measure the distance and velocity of objects relative to ego-vehicle enabling vehicles to conduct vehicle location and target recognition for realizing autonomous driving, which are conducive to vehicle collision avoidance, and can further improve the traffic efficiency in road networks.

2.2.1 GPS-based Techniques

GPS-based mobile devices (e.g., mobile phones and vehicle navigation systems) have attracted considerable attention owing to their continuously reduced costs and improved positioning accuracy [58]. These technologies provide rich traffic information for traffic condition estimation and congestion detection, such as precise vehicle positions and real-time travel speeds, accelerations, and travel directions. Herrera et al. [26] conducted a large-scale field experiment based on GPS-based probe vehicles and inductive loop detectors separately to evaluate the traffic monitoring feasibility and accuracy of GPS-enabled mobile phones in collecting traffic information. The results showed that the GPS-based approach can achieve realistic speed estimation and congestion detection. Based on GPS traces from probe vehicles, Yoon et al. in [25] presented speed thresholds of each road segment to categorize traffic on road segments as congested or free-flowing. In [59], Castro, Zhang, and Li calculated road traffic densities and predicted traffic conditions based on taxi GPS traces from Hangzhou, China. The experimental results validated the accuracy of the method in traffic congestion detection and prediction.

2.2.2 Mobile-phone-based Techniques

With the prevalence of mobile phones and the potential of collecting traffic data on vehicles over a large coverage area without requiring expensive infrastructure, mobile phones are being increasingly applied as traffic probes [60, 27, 35, 61]. Caceres et al. in [27] estimated road traffic flows by counting the vehicles switching from one cell to another. Experimental results showed that the reasonably accurate traffic flow estimates were achieved using the proposed method by comparison with vehicle volume measurements collected from loop detectors. In [35], Lv et al. estimated congestion degrees for road segments based on cellular signals and on-board accelerometers in mobile phones. Experimental results indicated the effectiveness and energy efficiency of the proposed system in vehicle monitoring and congestion estimation. By monitoring the period that a vehicle with mobile devices remained connected to a base station until it joined another one, Thajchayapong et al. presented a mobile-phone-based method to identify congested areas [61]. The field test results showed that the proposed congestion identification method based on cell dwell time agreed very well with that obtained from video records, which indicated the effectiveness of the method in congestion detection.

2.2.3 Radar/Lidar-based Techniques

The sensors in autonomous vehicles, such as radar and lidar, are also leveraged to collect data from road networks, which is an important prerequisite for the safe driving of autonomous vehicles, and can be further uti-
lized to detect road conditions. The measurement of radars mainly includes distance, relative velocity, and direction [62, 63]. The distance is estimated according to the round-trip time of electromagnetic waves from a radar via a target back to the radar. The relative velocity is calculated based on the Doppler effect and the direction is estimated by using the antenna array in radars. Danzer et al. [64] leveraged sparse radar data from autonomous vehicles to detect 2D objects in road networks. Experiment results indicated that the proposed strategy can achieve a higher classification accuracy than the other algorithms. Akita and Mita [65] utilized the Long Short Term Memory (LSTM) strategy to track and classify the target objects of the road environment, which indicated the potential to be applied for road condition detection in urban traffic networks.

Lidars can be used to detect target distance and create a 3D point cloud map around vehicles. Most existing techniques for target distance detection using lidars are based on the pulse measurement and phase shift measurement [66]. Pulse measurement utilizes the round-trip time of a laser pulse to and from a target to calculate the distance. Phase measurement measures the distance based on the phase difference between the emitted and returned laser pulses. Hata and Wolf [67] utilized a multilayer lidar to detect road environment features for vehicle localization to enable the obstacle detection and navigation systems for autonomous vehicles. Wang et al. [68] studied the pedestrian recognition and tracking problem for autonomous vehicles with the utilization of 3D lidar, which can promote the real-time path planning of autonomous vehicles. Jimenez and Naranjo [69] developed algorithms to enhance the obstacle detection and classification abilities using a lasercanner, which can benefit vehicle collision avoidance and further relieve the potential incident-induced congestion in road networks.

2.3 Vehicular Networks

In recent years, transportation researchers have shifted their focus to more advanced sensing technology, i.e., vehicular networks, because of the potential of vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication in acquiring real-time traffic information [70–72]. In VANETs, each vehicle could monitor its neighboring traffic conditions by on-board sensor units and exchange the detected data via V2V and V2I communication [73]. Then, each vehicle can estimate road conditions by computing the average travel speed, traffic density, and road travel times. In [74], Younes and Boukerche detected and analyzed congested road segments with reasonable accuracy and efficiency by using a cooperative vehicular communication protocol. Researchers in [75] developed an event-driven architecture to collect and process VANET messages for traffic congestion level detection. Simulation results suggested that the proposed mechanism can detect traffic congestion with short delay and provide detailed congestion information (e.g., congested lanes and queue length). In [76], Mallah et al. presented an approach for road congestion classification in urban road networks using VANETs, and the performance evaluation results showed the accuracy of the framework in classifying urban congestion.

2.4 Social Networks

With the rapid growth of mobile communication devices, social networks have become widespread recently, providing a modern form for real-time traffic information exchange. People can share ideas, express their opinions, and report real-life events based on social platforms such as Twitter, Weibo, and WeChat, which always contribute a large amount of valuable information for traffic condition detection (e.g., road congestion and accidents) [7]. In this case, many works have recently employed social media messages in traffic congestion detection. For example, Wanichayapong et al. [30] utilized syntactic analysis to extract congestion-related information from Twitter and classified them into traffic events and locations for the convenience of congestion reports. In [31], Andrea et al. developed a road condition surveillance system by extracting traffic information from tweets to identify real-time traffic jams and accidents. Experimental results indicated that the system distinguished traffic-related and unrelated information with 95.75% accuracy and extracted traffic events (e.g., congestion or crashes) with an accuracy of 88.89%.

III. TRAFFIC SIGNAL CONTROL
Traffic signal control has become widespread in almost all areas of the world and is generally acknowledged to be the most crucial and credible traffic control strategy, particularly in large cities. By setting time intervals for vehicles with traffic rights at each intersection, optimizing traffic signals can reduce congestion levels and vehicle travel time. The key parameters for traffic signal control usually involve phase, split, cycle length, and offset. A phase defines a combination of one or more traffic actions that receive priority at the same time during the signaling interval. The split is the total time allotted to each phase of a traffic light cycle. Cycle length is the time during the complete period of a traffic light cycle. The offset refers to the time delay between the start of green times at successive intersections. The explanation of a four-phase cycle at intersections is shown in Figure 3 to elaborate the four control variables.

There have already been numerous studies devoted to traffic light control strategies for mitigating road congestion and promoting traffic efficiency in urban areas. Generally speaking, traffic signal control can be categorized as fixed-time and traffic-responsive control strategies. Fixed-time control strategies are popular in most current urban traffic networks because of the easy implementation and low management cost. Moreover, fixed-time control strategies are suitable for regular traffic conditions with under-saturated traffic flows by employing the pre-determined cycle and split based on historical traffic demand. However, traffic systems are always dynamic and often non-predictable, resulting in the inefficiencies of fixed-time strategies. With the development of advanced sensing techniques, traffic-responsive control strategies have gradually attracted more attention and have improved on the limitation of fixed-time control strategies. Specifically, the splits, cycle lengths, and offsets for responsive strategies can be adjusted in real time according to the monitored road information, which makes traffic signals better able to adapt to real-time road conditions, reduce travel delays, and improve the efficiency of urban traffic networks.

3.1 Fixed-time Strategies

Fixed-time strategies with a pre-defined split and cycle length are suitably utilized at intersections with comparatively stable traffic flows. The earlier works on fixed-time strategies were applied to isolated intersections, which determine the optimal phase, split, and cycle to optimize travel delays and improve road capacity. Webster [77] established a traffic signal control strategy to minimize average vehicle delays that paved the way for the currently used traffic signal control strategies. Well-known examples of fixed-time traffic signal control at isolated intersections are SIGCAP and SIGSET proposed in [78] and [79], respectively. The former is suitable for intersections with heavy traffic demand to maximize the capacity of intersections and avoid over-saturation, while the latter is generally applied under unsaturated conditions to minimize the total intersection delays. To optimize earlier-generation fixed-time signal control for an isolated intersection, the strategies have been evolved to achieve traffic signal coordinated control in road networks. MAXBAND [80] developed a signal coordination system to maximize the bandwidth of the green wave band for vehicles by optimizing the offsets between successive intersections on bidirectional arteries. Then, many variants and extensions were developed, taking more new parameters into consideration, such as queue clearance time and left-turn movements [81]. Furthermore, offline software was also introduced in traffic signal control, providing the optimal signal arrangement for all intersections in traffic networks. TRANSYT [8] is perhaps the most well-known instance, which prepares a great quantity of fixed-time signal strategies for different times of day and chooses the most appropriate plan in real time according to the corresponding road conditions. The first on-site implementation of the signal control strategy evolved by TRANSYT showed an average travel time savings of 16% over road networks [82].

Although these fixed-time strategies cannot respond to real-time traffic conditions in a timely manner, es-
especially when some special events or traffic incidents occur, these strategies are still quite efficient in re-
ducing travel delays and mitigating urban traffic con-
gestion. In recent years, various sensing techniques,
such as loop detectors, have been more widely de-
ployed in large cities worldwide. Traffic-responsive
control strategies gradually indicate their adaptability
to changing traffic conditions, which provides more
potential methods for traffic signal control to better fit
real traffic networks.

3.2 Traffic-responsive Strategies

During the past few decades, driven by advanced
sensing and communication techniques, an increasing
number of responsive signal control strategies have
been presented that automatically respond to the pre-
vailing traffic conditions and indicate their greater
control efficiency than fixed-time strategies [83]. A
number of strategies, such as SCOOT and SCATS,
have been extensively applied in many cities around
the World and have achieved reduced travel delays and
road congestion. An illustration of the sensing and
communication signal control system is given in Fig-
ure 4. When vehicles arrive at a traffic signal, their
identification numbers, locations, arrival times, and
travel speeds can be collected using sensing technolo-
gies and V2V communications to estimate the real-
time traffic density. Then, the aggregated informa-
tion is disseminated to traffic signal controllers via
V2I communications and controllers in adjacent inter-
sections that can also communicate with each other
through wireless communications to share a wider
range of traffic information. Specifically, according
to the control scale of signal control systems, traffic-
responsive control strategies can be categorized as iso-
lated signal control, coordinated signal control, and
network-wide signal control strategies. The typical
strategies and their application conditions of these
three categories of traffic-responsive control strategies
are now presented.

3.2.1 Isolated Signal Control

Responsive isolated signal control strategies are appli-
cable to single intersections by adjusting signal phase,
split, and cycle length according to real-time traffic
conditions at each intersection. Chang and Park [84]
developed a signal control algorithm to enable unob-
strusted traffic at each intersection. The algorithm cal-
culated traffic volume and congestion degree on each
lane using V2V communications to decide the cycle
length and split for the next cycle. The algorithm
was verified under a single-intersection model and
showed its effectiveness in minimizing average vehi-
cle waiting times and queue lengths at intersections
compared with random control and best-first control
systems [85]. Pandit et al. [86] developed a signal
control method in which VANETs were employed to
detect average road speed, calculate the vehicle distri-
bution in networks, and then optimize the schedule of
traffic signals at each intersection. They formulated
the vehicular signal control issue as a job scheduling
problem and in which each job was regarded as a pla-
toon of vehicles passing through the same intersection.
Based on the wireless communications between vehi-
cles and roadside infrastructures, researchers in [87]
developed algorithms to schedule vehicle queues at in-
tersections on the basis of their journey lengths. Simu-
lation results showed that the proposed algorithms can
decrease congestion at intersections by 80% compared
to static signal control strategies.

3.2.2 Coordinated Signal Control

In addition to signal phase, split, and cycle length, co-
ordinated traffic signal control strategies consider the
offset between successive intersections to relieve con-
gestion at an urban zone or an entire road network
with many intersections. SCOOT is often consid-
ered a traffic-responsive version of TRANSYT, which
makes adjustments of splits, cycle lengths, and offsets
with a small-step increment on the basis of the mon-
tored traffic volume and occupancy from upstream
road segments. SCATS is another prevalent traffic sig-
nal control strategy applied in many cities, such as
Sydney, Melbourne, and Shanghai, and is deployed at more than 42,000 intersections in approximately 154 metropolises [88]. It uses automatic plan selection from a library in response to the real-time road conditions and adjusts signal parameters (e.g., cycle length, split, and offset) for each intersection in road networks [10, 89]. Specifically, at each intersection, the split and cycle length are determined depending on vehicle presence on each lane from inductive detective loops. Additionally, by setting a pre-determined offset during the start time of a particular phase, traffic signals for several successive nodes can be adjusted along a given path and a green wave can be generated. The self-organizing traffic lights (SOTL) system was proposed according to the principle that each intersection selects a phase with the highest demand. There is no direct coordination between the traffic signals at intersections in this system; however, indirect coordination is provided using the communication about the traffic arrivals and departures between neighboring intersections [90, 91]. Many other works based on back-pressure are also presented as dedicated to urban traffic-responsive signal control [92, 93]. These systems are constructed and implemented in a distributed manner with pre-determined or adaptive routes and choose the phase according to local information, such as queue length, on each road segment of an intersection. Simulation and analytical results show that maximum network throughput can be achieved by employing a back-pressure traffic signal control system, and better performance can be provided compared to those provided by fixed-time strategies and SCATS. However, these aforementioned strategies are usually effective when traffic conditions in a network are under-saturated; the performance of these strategies was reported to deteriorate in cases of saturated traffic conditions [94]. To this end, a number of model-based responsive strategies, such as PRODYN [95] and RHODES [96], have been developed for traffic signal control and utilize dynamic programming to optimize traffic signals of urban networks based on a high-fidelity model [97]. A traffic-responsive urban control (TUC) [98–100] system has been deployed on road networks of five metropolises and has been shown to be economical and efficient in both simulations and field implementations, even under saturated traffic conditions. Specifically, cycle length and offset adjustments are employed every few minutes and the corresponding values for the next cycle are computed based on the real-time traffic flows of current road segments. Split is controlled in every cycle, which predicts traffic flows on road segments during the next cycle time using a store-and-forward model to achieve split control for each cycle. As the key step to conduct signal control, this paper briefly reviews the TUC split control method, which is presented as follows.

First, the dynamic system for traffic flows on a road network is represented by using a store-and-forward model, as shown in (1).

\[
x(k + 1) = x(k) + B\Delta g(k),
\]

where \(x\) is a vector of the traffic flows on road segments; \(B\) is constant matrix reflecting the network characteristics; \(\Delta g(k) = g(k) - g^N\); \(g\) is a vector of green times for each phase in a cycle and \(g^N\) comprises the corresponding constant nominal green time \(g_i^N\).

Then, to minimize the negative effects of over-saturation and the spillover of vehicle queues and take the maximum road capacity into consideration, a quadratic criterion that combines both traffic flows and green times can be written as

\[
j = \frac{1}{2} \sum_{k=0}^{\infty} (x^T(k)Qx(k) + \Delta g^T(k)R\Delta g(k)),
\]

where \(Q\) and \(R\) are non-negative diagonal matrices.

The elements of \(Q\) are determined as the inverse of road capacity on each road segment. In addition, the step length of the control response can be affected by a weight matrix \(R = \tau I\), where \(\tau\) can be calculated during trial and error to realize a robust signal control in the whole road network.

Finally, based on the linear model (1) and an appropriate quadratic objective, a Linear-Quadratic-Regulator (LQR) approach is performed to obtain an effective gain matrix and enables a closed-loop solution, as given in (3).

\[
g(k) = g^N - Lx(k),
\]

where \(L\) is a solution for signal splits in each cycle, while cycle lengths and offsets can be determined according to other algorithms [101]. In summary,
the optimization problem in TUC is of quadratic programming, thus TUC works efficiently in coordinating traffic signal control on large road networks and has achieved satisfactory achievements in various countries, especially in saturated road situations.

Furthermore, vehicular-network-based techniques for responsive signal control have also attracted much attention and enable additional capabilities, such as predicting the duration before vehicles arrive at intersections and cooperating neighboring traffic signals to find the best schedule. Younes and Boukerche [102] developed a coordinated signal control strategy based on vehicular networks to provide the optimal vehicle scheduling for entire traffic networks. Simulation results indicated that the strategy can increase the traffic fluency and throughput by 70% compared to random and separate traffic light schedules [103]. Liu et al. [104] proposed a distributed cooperative signal control system in which a cooperative algorithm was used to achieve cooperation among controllers at successive intersections. Simulation results demonstrated that the proposed control system utilizing vehicular networks outperformed adaptive signal control methods in terms of road capacities, queue lengths at intersections, and vehicle waiting times. In addition, simulation results also provided an interesting notion that collaboration mechanisms were promising for traffic fluency and congestion mitigation in metropolises. Moreover, by using different artificial-intelligence-based methods (e.g., fuzzy systems, neural networks, and reinforcement learning), many advanced signal control strategies have been proposed, due to their intrinsic advantages, as follows. 1) These strategies can provide the optimal signal control policy without consideration of system models, and 2) the computational efficiency can be guaranteed on the high-dimensional state space.

Gokulan and Srinivasan [105] developed a responsive signal control method considering a fuzzy multi-agent framework. Simulation results demonstrated its efficiency in alleviating congestion at successive intersections. Jin and Ma [106] presented a multi-agent framework for intelligent traffic signal control based on reinforcement learning. A case study based on three neighboring intersections in Stockholm was developed to illustrate the advantages of the proposed framework in vehicle travel delays, fuel consumption, and number of stops. Prashanth et al. [107] proposed a reinforcement-learning-based intelligent signal control strategy, and indicated its advantages in terms of delays and queue length compared with the fixed-time strategies and SOTL [90]. Liang et al. [108] developed a deep reinforcement learning model for signal control in which the duration change of phases were actions and the reward was the cumulative latency difference between two successive cycles. Simulations based on SUMO illustrated the high efficiency of the proposed model in reducing average delays at intersections.

3.2.3 Network-wide Signal Control

All of the aforementioned strategies utilize a computationally heavy controller and are limited for use at one intersection or a few at best, which leads to the fact that these strategies are infeasible for large-scale traffic signal control due to the extremely high dimension of the joint action space. Recently, with the successful application of reinforcement learning, several studies of network-wide traffic signal control have been proposed attributed to the enhanced learning ability of reinforcement learning on complex tasks [109]. In [110], Chu et al. proposed a multi-agent deep reinforcement learning algorithm for traffic signal control in an entire road network. Simulation results based on the road network of Monaco city demonstrated the effectiveness of the proposed signal control strategy in robustness, average queue length, and intersection delays for the entire road network. According to the real-time traffic data collected by the Internet of Vehicles, Zhou et al. [111] developed a de-centralized reinforcement learning strategy for signal control, in which the Internet of Vehicles was utilized to gather traffic data and then transmit the data to an aggregation point for network-wide signal control. Simulation results showed that, compared with the state-of-the-art methods, the proposed method can decrease the convergence time while achieving comparable traffic control performance for the entire road network. Chen et al. [112] presented a decentralized multi-agent reinforcement learning approach to conduct a large-scale traffic signal control. As shown in Figure 5, each local agent broadcasts its state observation and receives state observations of other neighboring agents. Then each local critic network approximates its own value function for signal control. Simulation results based on SUMO confirmed the effectiveness of the proposed
control strategy in average travel speed, travel delay and queue length over other algorithms.

IV. VEHICLE ROUTE GUIDANCE

Vehicle route guidance enables drivers to obtain individual instructions from the departure location to the destination based on travel costs, such as travel time and distance. A route choice mechanism is the key technique of vehicle navigation systems that allows utilization of traffic networks in a more organized way and serves as an effective approach to manage urban traffic congestion [113, 114]. With the help of Big Data and communication technologies, finer-grained traffic information, such as congestion, queue lengths, average travel speeds, traffic flows, and delays at intersections, can be obtained and disseminated in real time. It is reasonable to predict that most vehicles will be equipped with wireless communication devices in the foreseeable future that allows vehicles to communicate with neighboring vehicles and roadside infrastructures [115]. In that event, dynamic vehicle routing based on constantly updating traffic information becomes possible, enabling vehicles to obtain a comprehensive global insight into traffic networks and perform the appropriate re-routing decisions to avoid unexpected congestion. Generally, according to how the route guidance system may react in urban road networks to conduct optimum routing guidance for drivers, route guidance strategies can be classified as follows.

- Deterministic versus stochastic: whether en-route vehicles will react and re-route their pre-determined travel route from the origin to destination (OD) considering the change of traffic conditions (e.g., traffic incidents and special events);

- Reactive versus predictive: whether the route guidance strategies consider the real-time traffic information only or the combination of the current road conditions and their prediction in the future;

- Centralized versus decentralized: whether a central system provides all route selections for all vehicles or individual vehicle needs to make its own path guidance decision.

4.1 Deterministic Versus Stochastic Strategies

Route guidance strategies can be divided into deterministic and stochastic strategies according to whether the travel costs of routes between different OD pairs are deterministic or stochastic and whether vehicles will change their pre-determined travel route in response to the unexpected road events, such as incidents and road construction.

Deterministic strategies determine and maintain paths from origins to destinations to achieve a more efficient route for vehicles without consideration of the temporal dependence and stochastic nature of traffic networks. Most of these strategies are proposed using shortest-path algorithms [116, 117], which provide paths between OD pairs for vehicles to minimize the sum of travel costs. The costs are usually defined based on vehicle travel time, which are determined at the start of journeys and without regard to the large variance of travel time in a stochastic traffic condition. A common algorithm is proposed in [116] via a Dijkstra algorithm to search the shortest paths between two nodes [116] or from a single origin as the “source” to many other destinations. Many variants have been developed based on a Dijkstra algorithm [118, 119]. In [118], Wagner and Willhalm outlined several latest extensions of a Dijkstra algorithm, which improved the computational speed of the Dijkstra algorithm and classified them into three classical techniques: bidirectional search (searching paths from source and destination simultaneously) [120], goal-directed search or A* algorithm (modifying the priority of active nodes to push towards the target) [119], and hierarchical methods (using a hierarchical architecture and considering a relatively small subgraph of the “hierarchical graph” to search the shortest path). Many other works on multiple shortest paths between an OD pair have also been proposed to provide a deterministic route guidance for vehicles; these works
take user preferences and alternative path constraints into consideration [121, 122] to produce more efficient navigation applications.

In addition to the development of deterministic route guidance strategies, it is important to realize that the shortest path does not always ensure a suitable and reliable travel time for vehicle navigation [123]. Owing to many uncertain factors in road networks, such as unexpected roadwork and incidents, traffic networks are neither static nor perfectly deterministic and, consequently, the shortest path may exhibit a large variance [124], which may lead to increased travel costs for an intuitively shorter route. Moreover, with the development of low cost and reliable sensing techniques (e.g., inductive loop detectors and probe vehicles), real-time traffic information becomes available gradually, which makes the dynamic routing strategies more feasible in modern urban traffic management systems. Hence, much attention has been paid to develop stochastic route guidance strategies to cater to such uncertainties in transportation systems [125–130].

In stochastic route guidance strategies, the optimal paths of vehicles may vary according to the uncertain nature of road traffic conditions. This uncertain nature is typically manifested in two aspects: temporal dependence (traffic conditions in the same roads are different in different times of the day) and randomness (there are many unexpected non-recurrent congestion events due to incidents or extreme weather causing difficulties in estimating travel times precisely) [125]. A number of studies of vehicle routing guidance that consider only the temporal dependence are available [126, 127]. Haghani et al. [126] presented a temporal route guidance strategy considering varying road travel times and compared vehicle travel times of the proposed strategies with those of static route planning. The results suggested that when the uncertainties of the travel time information increase, the dynamic routing strategy became increasingly superior to a static strategy. Lorini et al. [127] proposed a dynamic vehicle routing strategy considering temporal dependence and dynamic perturbations. The experimental results showed the advantages of the proposed strategy in travel time variance under the dynamic perturbation compared with static route strategies. In [128], Hall first considered both temporal dependence and stochastic natures of traffic networks to make routing choices for vehicles. Then, many studies were developed for temporally dependent and stochastic vehicle optimal path problems [123, 124, 129, 130]. Nie et al. [129] proposed a shortest-path strategy that allowed travelers to better plan their trips during random travel times with the prescribed vehicle arrival times. Researchers in [130] developed a stochastic route guidance model considering temporally dependent road travel times in the face of road congestion. Experiments were conducted based on the Singapore traffic network. Numerical results indicated that the presented dynamic strategy outperformed static routing guidance strategies in total travel time and routing efficiency.

4.2 Reactive Versus Predictive Strategies

Considering whether a route choice algorithm is reactive or predictive, vehicle navigation strategies can also be categorized as reactive or predictive. Reactive route choice strategies are only based on current conditions of a road network without consideration of future traffic conditions, while predictive route choice strategies utilize a traffic prediction model to forecast the road conditions in the near future and provide prospective path planning for vehicles. This classification is meaningful because it indicates the robustness and complexity of road guidance strategies. Reactive strategies are less complex compared with predictive approaches, while predictive strategies can provide greater robustness to congestion and incidents.

By deploying a road network traffic monitoring infrastructure, such as cameras and inductive loop detectors, the network-based service provides drivers with a current view of road conditions and “reacts” to determine the route selected [131]. In this circumstance, reactive route guidance, known as feedback routing, has been widely applied in route choice systems with massive real-time traffic information [132–137]. Furthermore, from an industry perspective, Google [132] and TomTom [133] have utilized monitored real-time road information to calculate the shortest routes for vehicles. Nericell [134] and CarTel [135] utilized probe vehicle data monitored by onboard devices to detect road conditions and estimate travel times for road segments in a network. In academia, Gao and Chabini [136] provided various approximations for paths with minimum travel costs, wherein a routing policy was
decided by drivers as the trip progresses according to the current travel times. In [137], Chen et al. proposed a real-time vehicle navigation algorithm based on wireless sensor networks that provided several real-time route choices considering drivers’ diverse navigation requirements. Attributed to the development of vehicular networks, Vondra, Becvar, and Mach [138] developed a route choice strategy with consideration of traveling time and throughput via both cellular and vehicular networks. Chai et al. [139] combined a route choice method through V2V and V2I communications that allowed travelers to make route choice decisions at each intersection according to real-time road information. Simulation results indicated the dynamic control technique could decrease average queue lengths at intersections and improve flow fluency for entire road networks. Zhang et al. [140] proposed a joint traffic-communication optimization strategy to transmit rerouting information via V2I communications for vehicles traversing the road segment with an incident. As illustrated in Figure 6, the Mobile Edge Computing (MEC) server is deployed at the roadside of a road segment and the MEC server can disseminate the rerouting traffic information with consideration of vehicle rerouting travel costs and communication costs. Simulation results demonstrated that the proposed strategy can achieve minimum total cost of vehicle rerouting and information transmission comparing with existing methods.

Leveraging the growth of sensing techniques, as introduced in Section II, massive traffic information is accurately and easily monitored in real time. In this case, the evolution process of traffic flows can be analyzed and discovered when these data can be utilized efficiently [141, 142]. As an important issue for intelligent transportation systems, many short-term traffic prediction models have been developed, such as the auto-regressive integrated moving average (ARIMA) model and its variants [40, 143, 144], Kalman-filter methods [145, 146], neural networks [147–149], and fuzzy techniques [150], which provide the shortest path in anticipation of future road traffic conditions to avoid potential traffic congestion. Predictive routing strategies, known as proactive strategies, have attracted great attention recently [131, 151, 152]. Pan et al. [131] developed several route choice strategies for drivers to proactively compute tailored rerouting guidance when congestion was predicted on their routes. Simulation results indicated that the route choice methods were valuable for congestion mitigation and can better cater to the spatio-temporal characteristics of road conditions than existing approaches. Claes et al. [152] proposed an anticipatory vehicle routing approach to avoid traffic jams that predicted road conditions in the near future to detect possible congestion and allowed vehicles to re-route. Simulation results suggested that the forecast vehicle routing method not only allowed vehicles to avoid congestion, but also prevented them from contributing to congestion as well. Wang et al. [153] developed a vehicle re-routing system by using V2I communications, which assisted drivers to make optimal route choices for congestion avoidance. Specifically, when an unpredictable traffic congestion was identified, an intelligent traffic light at each intersection was utilized to gather traffic information about real-time road conditions and vehicle destinations and provide the optimal routes for vehicles. Simulations showed that the proposed system could increase travel time reliability by 65.42% and reduce average travel times by 38.02% compared with the fastest and shortest rerouting methods [153]. Acting as a powerful location and navigation map, the high definition (HD) map can provide detailed road information around and further around the corner to support predictive cruise control [154]. In [155], Jian et al. proposed a local motion planning and obstacle avoidance method using the HD map. Simulations based on an autonomous driving experimental platform verified that the proposed method can improve the safety and comfort of vehicle motion planning. Takeuchi et al. [156] presented a blind area traffic prediction method based on the HD map and 3D lidar. Experiment results using the actual vehi-
cle with lidar in urban road networks demonstrated the effectiveness of the proposed method in safe velocity prediction compared with existing methods. Chu et al. [157] developed a predictive cruise control strategy based on the HD map information. Evaluation via simulation and experimental tests illustrated the proposed strategy can reduce 8.73% fuel consumption compared with the cruise control system without HD map information.

4.3 Centralized Versus Decentralized Strategies

Route choice in road networks can be computed by a centralized system that controls the state of the entire road network or by a distributed system that provides optimization from the viewpoint of individual drivers. The centralized routing scheme allows a reliable and robust system at the network level in which individual drivers may suffer because the users’ optimal routing may not always be consistent with that for the entire road network. However, the decentralized route choice method provides optimality concerns for individual vehicles and is not necessarily robust from a network perspective. Meanwhile, the centralized strategy has benefits in terms of area coverage, road coverage, route update interval, and computational complexity.

A centralized route choice strategy determines paths for all vehicles in a road network according to the collected traffic data from vehicles and road infrastructure using a centralized system that computes the new or re-routed path for a particular vehicle to minimize average traveling times at the network level and even relieve possible traffic congestion in traffic networks. Several works on shortest-path problems have been proposed using the centralized architecture [158–161]. Kanoh and Hara [158] formulated the shortest-path problem as a multi-objective optimization problem and utilized a centralized algorithm for individual vehicle navigation. In [161], Yamashita, Izumi and Kurumatani investigated a vehicle navigation mechanism using a centralized approach to reduce traffic congestion, in which drivers informed the route information server of their future routes and the server calculated prospective traffic conditions based on the collected routes, and then sent this information to the drivers. Based on vehicular social networks, Lin et al. presented a vehicle route selection algorithm in [162] in which the social correlations among vehicles were considered to categorize vehicles as different classifications and then calculated the optimal routes for each group of vehicles. Simulation results demonstrated the algorithm could achieve efficient traffic flow control and relieve traffic congestion in road networks.

Despite achieving better network throughput than decentralized systems, there are still two inherent problems with centralized systems. First, central servers must conduct intensive calculations to receive location updates and send the re-routing paths for vehicles, which makes centralized systems infeasible for large areas with a large number of vehicles. Second, a centralized approach requires real-time vehicle locations and a high compliance rate to achieve the effectiveness of the guidance systems, which will lead to major privacy concerns for users since a sequence of location can reveal drivers’ identities [164, 165]. Moreover, since transportation systems are often geographically distributed, as one of the powerful technologies for large-scale distributed systems, agent-based methods have emerged dramatically in recent years, indicating their great suitability for road traffic detection [166]. In this case, many decentralized systems have been developed to study vehicle route choice problems [152, 167–169]. Deflorio [167] proposed a decentralized and reactive route guidance system, which can achieve dynamic user equilibrium in a road network. Simulation results showed a travel time reduction as the penetration rate of the system increased. Tatamir and Rothkrantz [168] presented a hierarchical routing system to provide drivers with dynamic routing in which a hierarchy algorithm was introduced to divide traffic networks into multiple less complex networks. The algorithm showed high adap-
tivity to the dynamic changes caused by the special events in road networks, such as incidents and concerts. In [169], Wedde and Senge presented a decentralized vehicle routing guidance approach based on distributed vehicle-to-infrastructure architecture that provided a higher update frequency of traffic data information, even for large regions. Simulation results demonstrated that the proposed system outperformed dynamic shortest-path methods in data update frequency and average vehicle travel times.

V. SIGNAL CONTROL AND ROUTE CHOICE COMBINED TECHNIQUES

Route choice and traffic signal control often interact with each other [170]. As shown in Figure 7, non-equilibrium traffic flows on different routes between the same OD pair may cause the traffic control policy to change the green-time proportion for balancing the travel costs on different routes (travel costs often refer to the travel times). Such changes, in turn, give rise to fluctuations in route travel costs, which further causes vehicles to change routes for lower travel costs [170, 171], and the traffic network will be stable if and only if route choice and traffic signal control can reach equilibrium simultaneously [172, 173]. Based on this notion, researchers have developed many traffic control policies considering both traffic signal control and vehicle route choice to relieve congestion and enhance road capacity in traffic networks.

Smith presented a P0 control policy [173] and proved that a Wardrop equilibrium can be found by taking traffic signal control into consideration, which can achieve the maximum network throughput [170, 171]. Xiao and Lo [174] formulated a combined dynamical traffic system that encompassed both route choice and traffic signal control and investigated the interactions between adaptive traffic control and day-to-day route choice adjustments of travelers. In [113], Cao et al. developed a pheromone-based system to relieve road congestion that unified both signal control and dynamic vehicle re-routing. Simulation results showed that the system with combined vehicle re-routing and signal control outperformed other approaches that only considered vehicle re-routing or signal control strategy in traffic congestion and fuel consumption. In [175], a hybrid system that incorporated flow divergence and signal control in a general network was investigated. The proposed control policy and theoretical results derived in [175] can be utilized to boost road capacity and mitigate congestion in traffic networks. Moreover, with joint consideration of the route choice method, a novel traffic signal control policy was proposed in [163] to mitigate the negative effects of incident-induced congestion, which exhibited a more rapid recovery rate from an incident and improved the stability of a road network system experiencing incidents. The condition for the existence of new equilibrium after an incident was given in [163] and the stability of a road network system was also proved by using Lyapunov stability theorem. The proposed strategy was tested on a one-OD three-route road network and numerical results illustrated the effectiveness of the control policy.

VI. DISCUSSION

First, in this section, challenges in traffic control and management are discussed. Then, open data sources are introduced in subsection 6.2. Finally, future directions are also given.

6.1 Challenges

Although the existing research on congestion mitigation have made great achievements in ITSs, there are still many challenges that have not been fully addressed. The main open challenges of urban congestion control in ITSs are the following.

6.1.1 Public Dataset

A large variety of transportation applications have been studied based on available datasets, such as congestion detection, traffic signal control, and route guidance. However, much difficulty is caused by conducting direct comparisons between the proposed methods and algorithms due to lack of a common framework. One possible reason for this is the diversity of road topology, traffic conditions, and even local traffic regulations around the World. It would be beneficial for researchers of traffic management to determine several unified objectives for evaluation and comparison among different methods as achieved in visual object classes by PASCAL [176].
6.1.2 Data Collection

Owing to complicated road structures and rapid movement of vehicles in urban scenarios, data collected in ITSs might be inaccurate, incomplete, or unreliable in particular locations (e.g., tunnels) and at certain times (e.g., at night and in extreme weather) [177]. It is a possible way to deal with the challenges of investing new sensor technologies to improve the data collection capability. Specifically, based on continuously developing Internet of Things (IoT) techniques, more advanced sensing techniques can be invented that would effectively improve the automation and quality of traffic data collection.

6.1.3 Data Processing

Traffic data are a necessity for congestion-related applications, including traffic state recognition, traffic control, and vehicle re-routing. However, the existing traffic data are collected from diverse sources with different formats and should also be compared with historical data, which poses a significant challenge for data centers to process the massive amount of data in a short time. Many distributed data processing architectures have been presented based on cloud, fog, and edge computing [178–180], which offloaded a large portion of computation at the roadside infrastructures and vehicles, and thus, provided satisfactory solutions to achieving real-time data processing.

6.1.4 Security and Privacy

Security and privacy are enduring problems and have drawn widespread concern within the transportation and mobility domains. Privacy is possibly the most serious issue confronted by transportation authorities, which arises due to the conflicting requirements of opening and harnessing data to and for the public to improve transportation system performance and protect personal information from leakage [14]. For this reason, security and privacy protection are crucial for emerging technologies including vehicular networks and Big Data, and transportation administrators also face urgent challenges regarding private data collection. Surprisingly, there is a scarcity of references dedicated to guaranteeing the integrity, confidentiality, and security for traffic data in ITSs [181]. There is no doubt that overcoming these challenges with appropriate techniques will validly guarantee the utilization of traffic data and ensure the rights of traffic participants in compliance with privacy recommendations, regulations, and laws.

6.2 Open Data Sources

As mentioned above, open data sources are important for research in transportation and contribute valuable information and knowledge to ITSs. Unfortunately, most existing works utilized proprietary traffic data, which are rarely available online. This fact was raised in [182], in which guidance was provided for transportation administrators, including opening and maintaining data, overcoming potential barriers for their implementation, and improving communications between transportation organizations and their customers.

Here, a number of open data sources currently used in transportation systems are introduced. In the U.S., 228 transit agencies share their schedule data to Google [183] (more than 475 around the World), which publishes the data as part of its Google Maps service. A geo-spatial analytics database is shared [184] by the San Francisco Department of Public Health, which includes the monitoring, analysis, and evaluation of traffic spatial data and injury data to promote public health and equity. Another open geographic dataset for ITSs was exhibited in [185], which unveiled correlation patterns between geographic features and user preferences by using the public Yelp dataset. For video camera data, grayscale images of road intersections from long-distance and high-vantage perspective are available in [186] and a new three-dimensional dataset for visual traffic monitoring is provided in [187].
6.3 Future Research

6.3.1 5G and Beyond Communications

To strengthen the connectivity of vehicles, 5G and beyond technologies, enhanced by multiaccess edge computing, provide flexible and cost-effective solutions to support vehicle-to-Everything communications [188, 189]. Specifically, 5G and beyond technologies allow vehicles to receive a tremendous volume of real-time traffic data via broadcasting messages from the surrounding environment, which can provide effective information services about route guidance, collision warning and congestion avoidance. Furthermore, due to the advantages of massive device connectivity, high data rate and low latency, 5G and beyond communications technologies have the potential to be widely applied to enable a variety of advanced vehicle-to-Everything applications [190, 191], such as vehicle platoons and cooperative driving, which can promote a more convenient and safer driving environment, alleviate road congestion and improve traffic flows in road networks.

6.3.2 Big Data for Vehicular Networks

As the expansion of IoT and vehicular applications, massive amounts of data are generated from various sources with different formats, locations, and information contents, and have become an important concern for vehicular networks [192]. To address the intrinsic shortcomings of centralized control systems and offload their computational burden at local servers, cloud, fog, and edge computing based on Big Data analysis have been considered a promising way for the continuous evolution of vehicular networks, which have boosted the existing vehicular networks to the Big Data era [193]. Based on [194], a Big Data architecture for vehicular networks is presented in Figure 8 that is composed of two layers: the fog layer and edge layer. The architecture with combined vehicular networks and Big Data analysis is expected to have advantages over traditional vehicular networks in its latency, efficiency, coverage, and service scalability, which hence leads to the potential to be applied and deployed in real traffic networks and provides a flourishing future research direction for congestion mitigation.

6.3.3 Vehicle-road Cooperation Systems

By deploying a large amount of roadside IoT equipment, the vehicle-road cooperation system is considered an effective means of reducing traffic congestion and improving traffic safety [195], and provides a potential application scenario for the integration of enhanced mobile broadband (eMBB), massive machine type of communication (mMTC), and ultra-reliable and low latency communications (URLLC) users in 5G and beyond wireless communications. Specifically, by using information, sensor, and wireless communications and control technologies, vehicle-road cooperation systems can achieve real-time perception of vehicle and road conditions, safety status analysis, and emergency warnings. This strengthens the information exchange between vehicles and roads, and is used to optimize the operation and management of intelligent transportation systems, thereby forming a safe and efficient transportation system.

6.3.4 Connected and Autonomous Vehicles

Connected and autonomous vehicles (CAVs), which allow vehicles to receive non-line-of-sight road information from Vehicle-to-Everything communications, provide an opportunity for solving traffic problems and supporting sustainable development in urban road networks. In recent years, much attention has been paid to the research of CAV technologies. Traditional vehicle companies, e.g., BMW, Audi and General Motors, and Internet auto companies, e.g., Google, Uber and Tesla are participating in the manufacturing and research of CAVs to achieve enhanced driving safety, increased traffic throughput and reduced fuel consumption [36, 196]. Furthermore, because CAVs are appearing in road networks, and the heterogeneous traffic environment with mixed human drive vehicles and CAVs is expected to last a long time [197, 198], understanding the effects of CAVs on heterogeneous traffic environments is essential. In this case, establishing a novel traffic control mechanism that is friendly to both types of vehicles is still a prosperous research area to promote the efficiency and safety of transportation systems.

VII. CONCLUSIONS
In this work, a comprehensive literature review of traffic congestion mitigation strategies is conducted, with a specific focus on urban road networks. The literature survey covers the necessary pre-requisites for the implementation of congestion control, i.e., congestion detection. Then, various strategies for signal control, route choice, and their integration are elaborated and discussed. Furthermore, as the new trends for research in congestion mitigation, the architectures and applications of vehicular networks and Big Data are also introduced.

Challenges and the potential research trends regarding traffic congestion control may help guide future research in this area. The initial proposition of this survey may be purposefully used by academia and industry. Further extensive research studies for raising new issues and challenges in this topic are called for.

ACKNOWLEDGEMENT

This work was supported by the National Key R&D Program of China (2019YFB1600100), National Natural Science Foundation of China (U1801266), and the Youth Innovation Team of Shaanxi Universities.

References


[54] W. Ma, D. Xing, et al., “A wireless accelerometer-

Authorized licensed use limited to: XIDIAN UNIVERSITY. Downloaded on January 06,2022 at 03:06:39 UTC from IEEE Xplore. Restrictions apply.


[87] F. Ahmad, S. A. Mahmud, et al., “Shortest processing time scheduling to reduce traffic congestion in dense urban areas,” IEEE Transactions on Systems, Man, and Cybernet-


[183] [Online]. Available: www.citygoround.org


Biographies

Wenwei Yue received B.E. and the Ph.D. degrees in communication and information systems from Xidian University, Xi’an, China, in 2015 and 2020, respectively. He was also a visiting Ph.D. student in School of Computing and Communications, University of Technology Sydney, from 2017 to 2019. He is currently an Associate Professor with the State Key Laboratory of Integrated Services Networks, Xidian University. His research interests include intelligent transportation systems, vehicular networks, wireless sensor networks and big data.

Changle Li received the Ph.D. degree in Communications and Information System from Xidian University, China, in 2005. He conducted his postdoctoral research in Canada and the National Institute of Information and Communications Technology, Japan, respectively. He had been a Visiting Scholar with the University of Technology Sydney and is currently a Professor with the State Key Laboratory of Integrated Services Networks, Xidian University. His research interests include intelligent transportation systems, vehicular networks, mobile ad hoc networks, and wireless sensor networks.

Guoqiang Mao has published over 200 papers in international conferences and journals, which have been cited more than 9000 times. He is an editor of the IEEE Transactions on Intelligent Transportation Systems (since 2018), IEEE Transactions on Wireless Communications (2014–2019), IEEE Transactions on Vehicular Technology (since 2010) and received “Top Editor” award for outstanding contributions to the IEEE Transactions on Vehicular Technology in 2011, 2014 and 2015. He was a co-chair of IEEE Intelligent Transport Systems Society Technical Committee on Communication Networks. He has served as a chair, co-chair and TPC member in a number of international conferences. He is a Fellow of IET. His research interest includes intelligent transport systems, applied graph theory and its applications in telecommunications, Internet of Things, wireless sensor networks, wireless localization techniques and network modeling and performance analysis.

Nan Cheng received the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Waterloo in 2016, and B.E. degree and the M.S. degree from the Department of Electronics and Information Engineering, Tongji University, Shanghai, China, in 2009 and 2012, respectively. He worked as a Post-doctoral fellow with the Department of Electrical and Computer Engineering, University of Toronto, from 2017 to 2019. He is currently a professor with State Key Lab. of ISN and with School of Telecommunication Engineering, Xidian University, Shaanxi, China. His current research focuses on B5G/6G,
Hspace-air-ground integrated network, big data in vehicular networks, and self-driving system. His research interests also include performance analysis, MAC, opportunistic communication, and application of AI for vehicular networks.

Di Zhou received the B.E. degree in information engineering and the Ph.D. degree in communication and information systems from Xidian University, Xi’an, China, in 2013 and 2019, respectively. She was also a Visiting Ph.D. Student with the Department of Electrical and Computer Engineering, University of Houston, from 2017 to 2018. Since 2019, she has been with the Broadband Wireless Communications Laboratory, School of Telecommunications Engineering, Xidian University, where she currently holds a faculty postdoctoral position. Her research interests include intelligent transportation systems, routing, resource allocation, and mission planning in space information networks.