Optimization of an Emergency Response Vehicle’s Intra-Link Movement in Urban Transportation Networks Utilizing a Connected Vehicle Environment

Gaby Joe Hannoun

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirement for the degree of

Doctor of Philosophy
In
Civil Engineering

Kevin P. Heaslip, Co-Chair
Pamela M. Murray-Tuite, Co-Chair
Kathleen L. Hancock
Thidapat Chantem

June 24th, 2019
Blacksburg, Virginia

Keywords: intelligent transportation systems, connected vehicles, emergency services, emergency response vehicles, integer linear programming, vehicle routing
ABSTRACT

Congestion poses growing risks to emergency response vehicles (ERVs). These vehicles, striving to reach their destination as quickly as possible because people’s lives may depend on them, are often forced to make risky maneuvers to find a clear passage. In this research, an advanced intra-link emergency assistance system, that leverages the emerging technologies of the connected vehicle environment, is proposed. The proposed system assumes the presence of a centralized system that gathers/disseminates information from/to connected vehicles via vehicle-to-infrastructure (V2I) communications. The major contribution of this dissertation is the intra-link level support that is provided to ERVs as well as non-ERVs. The proposed system provides network-wide assistance as it also considers the routing of ERVs. The core of the system is an Integer Linear Program (ILP) that generates, based on location and speed data from connected vehicles that are downstream of the ERV, the fastest intra-link ERV movement. It specifies for each connected non-ERV a final assigned position that it can reach comfortably along the link. The ILP accommodates partial market penetration levels. In addition, a sequential ILP optimization approach is adopted to make it applicable on large transportation link segments. The complete emergency response assistance system consists of three modules (1) an ERV route generation module, (2) a criticality analysis module and (2) the sequential ILP optimization module (extension of the earlier approaches). The first module determines the ERV’s route (set of links) from its origin to its desired destination in the network. Based on this route, the criticality analysis module scans/filters the connected vehicles of interest and determines whether any of them should be provided with a warning/instruction message. As the ERV is moving downstream, new non-ERVs should be considered. So, the criticality analysis module is continuously scanning downstream traffic until the ERV reaches its destination. When a group of non-ERVs is identified by the criticality analysis module, the sequential ILP optimization module is activated. The criticality analysis module excludes non-ERVs diverging from the ERV route at intersections to avoid sending unnecessary requests. In addition, the sequential ILP optimization module ensures that the ERV occupies the most appropriate lane prior entering an intersection (i.e., lane depends on the intended movement). The system is evaluated using NetLogo, an agent-based modeling environment, under different combinations of market penetration and congestion levels. A minimum market penetration of 40% is recommended to guarantee improvements in ERV travel time. Benefits in terms of ERV travel time with an average reduction of 9.09% and in terms of vehicular interactions with an average reduction of 35.46%
and 81.38% for ERV/non-ERV and non-ERV/non-ERV interactions respectively are observed at 100% market penetration, when compared to the current practice where vehicles moving to the nearest edge.
OPTIMIZATION OF AN EMERGENCY RESPONSE VEHICLE’S INTRA-LINK MOVEMENT IN URBAN TRANSPORTATION NETWORKS UTILIZING A CONNECTED VEHICLE ENVIRONMENT

Gaby Joe Hannoun

GENERAL AUDIENCE ABSTRACT

Downstream vehicles detect an emergency response vehicle (ERV) through sirens and/or strobe lights. These traditional warning systems do not give any recommendation about how to react, leaving the drivers confused and often adopting unsafe behavior while trying to open a passage for the ERV. In this research, an advanced intra-link emergency assistance system, that leverages the emerging technologies of the connected vehicle environment, is proposed. The proposed system assumes the presence of a centralized system that gathers/disseminates information from/to connected vehicles via vehicle-to-infrastructure (V2I) communications. The major contribution of this dissertation is the intra-link level support provided to ERV as well as non-ERVs. The proposed system provides network-wide assistance as it also considers the routing of ERVs. The core of the system is a mathematical program - a set of equations and inequalities - that generates, based on location and speed data from connected vehicles that are downstream of the ERV, the fastest intra-link ERV movement. It specifies for each connected non-ERV a final assigned position that the vehicle can reach comfortably along the link. The system accommodates partial market penetration levels and is applicable on large transportation link segments with signalized intersections. The system consists of three modules (1) an ERV route generation module, (2) a criticality analysis module and (2) the sequential optimization module. The first module determines the ERV’s route (set of links) from the ERV’s origin to the desired destination in the network. Based on this selected route, the criticality analysis module scans/filters the connected vehicles of interest and determines whether any of them should be provided with a warning/instruction message. As the ERV is moving towards its destination, new non-ERVs should be notified. When a group of non-ERVs is identified by the criticality analysis module, a sequential optimization module is activated. The proposed system is evaluated using simulation under different combinations of market penetration and congestion levels. Benefits in terms of ERV travel time with an average reduction of 9.09% and in terms of vehicular interactions with an average reduction of 35.46% and 81.38% for ERV/non-ERV and non-ERV/non-ERV interactions respectively are observed at 100% market penetration, when compared to the current practice where vehicles moving to the nearest edge
Dedication

To my parents Elie and Fadia,
To my fiancé Johnny,
To my brother Joseph,
To my friends,

Thank you for your endless love and support,
Thank you for being a part of this journey and keeping up with my emotional rollercoaster,
I could not have done it without you.
Acknowledgements

My deepest and most sincere gratitude goes to my advisors Dr. Kevin Heaslip and Dr. Pamela Murray-Tuite for their support, patience and continuous encouragement throughout my journey. Thank you for sharing your knowledge and expertise and most importantly for teaching me valuable life lessons. For that, I can never repay you. Dr. Kevin Heaslip, I promise to pass on your positivity and trustworthiness and I hope to inspire others as you have inspired me. Dr. Pamela Murray-Tuite, your perseverance and meticulousness are exceptional and I look forward to follow your footsteps.

Thank you to my committee members, Dr. Thidapat Chantem and Dr. Kathleen Hancock for their time, constructive feedback and guidance. I greatly appreciate your efforts to get my dissertation this far. Working with you has been a privilege.

Thank you to my undergraduate advisor Dr. John El-Khoury who believed in me and encouraged me to pursue my graduate studies. This decision has changed my life and I will forever be grateful to you.

Thank you to my colleagues with whom I shared this memorable journey. Thank you for your friendship and for supporting me, each in your own unique way.
# Table of Contents

Chapter 1. INTRODUCTION ........................................................................................................... 1  
1. 1. Overall goal and research questions .............................................................................. 1  
1. 2. General approach ............................................................................................................. 2  
1. 3. Anticipated contributions .............................................................................................. 2  
1. 4. Dissertation outline ......................................................................................................... 3  

Chapter 2. BACKGROUND AND LITERATURE REVIEW ................................................................. 5  
2. 1. Overview of current ERV operations ............................................................................... 5  
2. 1. 1) Challenges experienced by ERVs ............................................................................... 5  
2. 1. 2) Currently deployed measures .................................................................................... 7  
2. 2. Overview of the Connected Vehicle Environment .......................................................... 14  
2. 2. 1) Applications at intersections ...................................................................................... 15  
2. 2. 2) Applications at links .................................................................................................. 19  
2. 2. 3) Other applications: accurate measurements ............................................................... 21  
2. 3. Connected vehicles applications assisting the ERVs movement ..................................... 22  
2. 3. 1) At intersections .......................................................................................................... 22  
2. 3. 2) At links ....................................................................................................................... 23  
2. 3. 3) Other emergency assistance systems ......................................................................... 24  

Chapter 3. FACILITATING EMERGENCY RESPONSE VEHICLES’ MOVEMENT THROUGH A ROAD SEGMENT IN A CONNECTED VEHICLE ENVIRONMENT ............................................... 31  
3. 1. Introduction .................................................................................................................... 31  
3. 2. Literature review .......................................................................................................... 32  
3. 3. Problem statement and assumptions .............................................................................. 33  
3. 3. 1) Problem Statement .................................................................................................. 33  
3. 3. 2) Roadway Discretization and Labeling ...................................................................... 33
List of Figures

Figure 3-1: Discretization of roadway network ................................................................. 33
Figure 3-2: Initial range, feasible stopping ranges, and assignment range ...................... 34
Figure 3-3: Variation of ERV speed in the AR with different ERV initial positions (per road type and ERV size). LC refers to “Lane Change”, so charts without “LC” indications represent a straight ERV path .......................................................................................................................... 43
Figure 3-4: Variation of ERV speed in the AR with different practices (Major Collector/Ambulance) ... 48

Figure 4-1: Link segment divided in IRs .............................................................................. 59
Figure 4-2: Sequential optimization flowchart ................................................................... 59
Figure 4-3: Non-ERVs j and j’ with corresponding FSR (when speed of j is higher than the one of j’) ... 67
Figure 4-4: FSR and AR overlap ............................................................................................ 69
Figure 4-5: Average computation time per IR for different numbers of IRs within the link and different levels of congestion ......................................................................................................................... 77
Figure 4-6: Groupings of non-ERVs in IRs ............................................................................ 78
Figure 4-7: Percentage of connected non-ERVs that adjusted their instructed final longitudinal position for different v/c ratios and market penetration levels ................................................................................................................. 80

Figure 5-1: Proposed system composed of three modules .................................................. 96
Figure 5-2: Data flow ........................................................................................................... 97
Figure 5-3: Initial ranges (IRs) and corresponding assignment ranges (ARs) ...................... 103
Figure 5-4: Different heading for different links upstream an intersection ...................... 104
Figure 5-5: X and Y axis for different headings .................................................................. 104
Figure 5-6: ERV route in a transportation network as pictured by the sequential ILP optimization module. The optimal ERV route is highlighted in (a). Figures (b) through (e) show how the links are virtually rotated for X and Y axis’ consistency and the intersection space is excluded ........................................... 105
Figure 5- 7: Non-ERVs inside intersection at detection are considered positioned on the stop line........106

Figure 5- 8: Network representation showing the link and node labels........................................118

Figure 5- 9: Percent reduction in ERV travel time for each market penetration level at a v/c ratio equal to 0.75 .................................................................................................................................121

Figure 5- 10: Percent reduction in ERV travel time for each market penetration level at a v/c ratio equal to 1.....................................................................................................................................122

Figure 5- 11: Percent reduction in ERV travel time for each market penetration level at a v/c ratio equal to 1.25 ........................................................................................................................................122

Figure 5- 12: Percent reduction in vehicular interactions for different market penetration levels ..........123

Figure D- 1: Sample time-space diagram of conflicting ERV and non-ERV ........................................126

Figure D- 2: Time-space diagram showing how ε is computed .........................................................128

Figure D- 3: Time-space diagram with different non-ERVs and showing the Δdadd1 .........................129

Figure D- 4: Time-space diagram with different non-ERVs and showing the Δdadd2 .......................130

Figure D- 5: Prohibit entry of new non-ERVs Δtne prior tdd.................................................................131
List of Tables

Table 2- 1: Current response times vs target response times ..........................................................7
Table 2- 2: Review of studies evaluating transition methods after ERV preemption .......................10
Table 2- 3: Review of studies evaluating transition methods in general ..........................................11
Table 2- 4: Adaptive traffic signal control using connected vehicles .............................................16

Table 3- 1: Variable notation .............................................................................................................35
Table 3- 2: Parameter notation ........................................................................................................35
Table 3- 3: Base case scenarios parameters and resulting ERV path description .........................41
Table 3- 4: Limiting scenarios input/output data .............................................................................45
Table 3- 5: Objective function weight analysis results .....................................................................46
Table 3- 6: Final speeds for local practice vs formulation ...............................................................47

Table 4- 1: Parameter notation ........................................................................................................62
Table 4- 2: Variable notation .............................................................................................................63
Table 4- 3: Computation times with and without rule 1 ................................................................67
Table 4- 4: Computation times (with rules vs without rules), where the initially inputted c value is 2 cells when no rules in used ......................................................................................70
Table 4- 5: Parameters used for scenarios in Test A .......................................................................77
Table 4- 6: Average computation time per AR per grouping shown in Figure 4- 4 .........................78
Table 4- 7: ERV travel time reduction due to our proposed system ................................................81

Table 5- 1: Parameter notation ........................................................................................................106
Table 5- 2: Variable notation ...........................................................................................................107
Table 5-3: Percent increase in impact (in terms of duration affected by the ERV passage) on non-ERVs at each market penetration level.

Table B-1: Estimated CT based on the IR longitudinal size and non-ERVs number.
Chapter 1. INTRODUCTION

First responders are required to answer a call and arrive as soon as possible at the emergency scene in which a situation requires immediate action. The Emergency Medical Services (EMS) vehicles crash data report an annual average of 4500 ambulance crashes between 1992 and 2011 in the United States [1]. Out of these crashes, an estimate of 65% resulted in property damages, 34% led to 2600 injured persons per year, and less than 1% led to an average of 33 fatalities. Regarding the firefighters’ vehicles crash data, vehicle collision is ranked as the second major reason behind on-duty fatalities of firefighters [2]. The number of firefighter fatalities caused by vehicle crashes constituted 14% of the total firefighters’ deaths in 2014 [2]. According to the U.S. Firefighter Injuries 2014 report, 550 out of a total of 14,910 collisions related to fire department vehicles resulted in firefighters injuries [3]. Vehicle crashes have been identified as the main cause of death of police officers with an increasing number of officers killed in motor vehicle crashes since the 1990s [4].

Developing a safe, timely, and effective emergency response system is indispensable to our communities. Safety is paramount since the emergency response system should assist people and often saves their lives but should not place them at greater risk. Emergency response vehicles (ERVs) have a unique design and characteristics. They navigate dangerously, especially when responding to an emergency. Unfortunately, roads are rarely designed for this type of use; also, it is very challenging to notify all the road users of an approaching ERV and to coordinate their corresponding actions to ensure the fastest ERV passage. Therefore, when traveling from origin to the emergency site, an ERV is subject to delays as well as to high risks of crashes that will hinder its movement and prevent it from reaching its destination quickly and safely. The existing emergency vehicle signal preemption concept only facilitates the passage of ERVs at a signalized intersection. Also, the traditional emergency warning systems such as the sirens and lights have only the capability to provide limited detection times of an approaching ERV; thus, they cannot be considered completely effective when it comes to facilitating the ERVs passage on transportation links. With advancements in wireless and communication technologies, the development of an advanced emergency assistance system is valuable to provide early detection of ERVs as well as assistance messages to downstream traffic to facilitate the overall passage of ERVs.

1.1. Overall goal and research questions

In this research, the overarching goal is to develop a semi-automated system that assists an ERV traveling from one origin to a destination in a network using connected vehicle technology. The proposed system first defines an optimal dynamic path from origin to destination. Then, it assists the ERV by optimizing its intra-link movement and determining the best ERV maneuvering actions to take when
traveling along each link in its path and by instructing the downstream connected non-ERVs to brake and stop at specific positions along the link before the arrival of the ERV.

Accordingly, the following research questions will be answered:

- What are the benefits (regarding ERV travel time improvement and vehicular interaction reduction) observed on transportation links that the proposed system can offer over the currently adopted practices fully relying on human response and traditional warning systems?
- Is the proposed system beneficial under different partial market penetrations (regarding ERV travel time improvement and vehicular interaction reduction)?

1.2. General approach

This approach primarily relies on a mathematical program inspired by initial work with nonlinear components in [5]. It is an integer linear program that applies to a transportation segment when the non-ERVs traveling on that segment is expected to conflict with the ERV path or hinder its movement. Based on the collected vehicle data of each connected non-ERV traveling on this link segment, the projected range where these non-ERVs can stop is identified along which the ILP generates the fastest ERV intra-link passage with the corresponding ERV maneuvers. Also, the ILP assigns each non-ERV to a specific location along the projected range which is communicated with each non-ERV in a specific instruction message. In conjunction with a vehicle routing algorithm to determine the optimal ERV route in the transportation network, the developed ILP is applied sequentially on the link segments (constituting the ERV route); hence, the intra-link movement on the complete ERV route from origin to destination is facilitated (i.e., the fastest intra-link movement is defined). This approach accounts for partial market penetration conditions by applying an estimation technique to predict the presence of unconnected non-ERVs between a pair of connected non-ERVs. Hence, the estimated non-ERVs will be virtually reserving a space where the actual unconnected non-ERVs can comfortably stop.

1.3. Anticipated contributions

The major contribution of this dissertation is to assist ERVs while traveling on transportation links. The core of this system is an integer linear program that is developed to optimize the intra-link movement on a link segment. This integer linear program is associated with several other components, to deliver a complete application on a network-wide basis. First, it is combined with a vehicle routing algorithm that determines the initial ERV route from the origin to destination on which the ILP should be executed. Then, based on the selected ERV route, the ILP will be applied to the link segments (constituting the selected route) sequentially, as the ERV is traveling. Another component of the proposed system identifies the time at which the ILP should be executed on a specific link in a way that allows the non-ERVs to stop at their
final instructed positions before the ERV arrival. With this advanced emergency assistance system, considerable improvements regarding ERV travel time is expected as the ILP’s objective is to maximize the ERV speeds. Plus, a reduction in vehicle interactions (ERV/non-ERV and non-ERV/non-ERV) is anticipated as the objective function also maximizes the free space around the ERV intra-link path. Furthermore, the positions of non-ERVs are determined in a way to ensure (1) minimum stopping distance with comfortable deceleration rates and (2) no conflict among non-ERVs as weaving is prohibited and passing is limited. Another contribution is the applicability of this system during the transition to a time when vehicles are all fully connected, making it useful in the short-term.

1.4. Dissertation outline

This dissertation will follow the manuscript format where published, submitted and in preparation for submission papers are replacing some of the standard chapters. The proposed outline of the next chapters is as follows:

- **Chapter 2. Background Information and Literature Review**
  This chapter first describes the current operations of ERVs on the roads along with the currently deployed countermeasures. Second, it discusses the connected vehicle environment and third the proposed systems introduced in the literature to assist ERVs using advanced wireless and communication technologies.

- **Chapter 3. Facilitating Emergency Response Vehicles’ Movement Through a Road Segment in a Connected Vehicle Environment**
  This chapter is a published paper in IEEE Transactions on ITS. It introduces an integer linear program that generates the fastest intra-link ERV movement while assigning each downstream non-ERV to a specific position along the link. This paper assumes a fully connected vehicle environment (100% market penetration).

- **Chapter 4. Sequential Optimization of an Emergency Response Vehicle’s Intra-Link Movement along Transportation Segments in the Connected Vehicle Environment**
  This chapter is a paper submitted to the journal of Transportation Research Part C: Emerging Technologies. It is an extension of the paper proposed in Chapter 3. The extended sequential optimization approach applies to larger transportation links and uses new information that becomes available as the ERV is moving forward. It also accommodates unconnected downstream non-ERVs (partial market penetration).

- **Chapter 5. Optimization of an Emergency Response Vehicle Movement from Origin to Destination in an Urban Transportation Network using Connected Vehicle Technologies**
This chapter is a paper in preparation for submission. It introduces a complete semi-automated system that assists the ERV while traveling in an urban transportation network. After identifying the optimal ERV route from one origin to a destination, a heuristic technique is executed to detect the downstream non-ERVs that should be provided with an instruction message. Accordingly, the sequential ILP optimization approach, discussed in Chapter 4, is used (1) to optimize the ERV intra-link movement along the transportation link segment where the detected non-ERVs can potentially stop and (2) to generate the non-ERVs’ messages. In this paper, the system is extended to accommodate the presence of intersections.

- **Chapter 6. Conclusion**
  
  This chapter summarizes the key findings of this dissertation and discusses the future works.

**Reference**


Chapter 2. BACKGROUND AND LITERATURE REVIEW

In this chapter, an overview of the current emergency vehicle operations is discussed followed by a description of the currently deployed measures that aim to facilitate the ERVs’ passage on the roads. Next, an overview of the connected vehicle environment is introduced along with its challenges, benefits, and most relevant applications. Furthermore, a review of the studies proposing new and advanced systems to assist emergency vehicles while relying on advanced wireless and communication technologies is discussed.

2.1. Overview of current ERV operations

The deployed measures for ERV support in transportation systems are limited to the emergency preemption at signalized intersections and traditional warning systems such as sirens and lights. Nevertheless, according to the crash reports, emergency vehicles operations are characterized by high risk of collisions [1]. Also, their response times often exceed the maximum acceptable thresholds. Thus, the current ERVs’ activity shows that the deployed measures are not sufficient.

2.1.1) Challenges experienced by ERVs

A crash involving an ERV can result in fatalities, injuries and property damages related to this specific crash. Secondary effects include the crash-involved ERV ceasing response to the original call and requiring the deployment of another ERV to the original call, delaying assistance. Furthermore, additional ERVs are likely to be needed at the ERV-involved collision, further straining limited resources. Further analysis of the crash data reports showed that ERV crashes also have a significant impact on the other-vehicle occupants and non-car occupants (pedestrian and pedalcyclist). For instance, the traffic safety facts for 2013 [2] revealed that the combined number of occupants of the other-vehicle and non-car occupants killed is notably greater than the number of ERV occupants killed for ambulances, fire trucks, and police vehicles. Also, the NHTSA’s Fatality Analysis Reporting System (FARS) 1992-2011 reported that the number of occupants of the other vehicles killed and injured in crashes involving an ambulance was 63% and 54% of the total persons killed and injured, respectively. It is important to define the causes behind the ERV’s crashes to identify the required countermeasures. A study focusing on the characteristics of ERV operations in the Washington D.C. region between 1997 and 2001 found that crashes along the U.S.1 in Fairfax County, Virginia occurred mostly at intersections (64%). The study revealed that angle type collisions had the highest frequency and were a result of a failure in maneuvering, conflict when crossing an intersection and making a left turn [3].

Similarly, Vrachnou [4] conducted a study on the U.S. highways and interstates in Northern Virginia between 1997 and 2001. The study showed that 75% of the ERV crashes at intersections in Fairfax County occurred at signalized intersections. Vrachnou concluded that the lack of sufficient warning signs
to alert vehicles with conflicting paths was the primary cause of those accidents. Ambulance crash data from NHTSA FAR and NASS GES collected between 1990 and 2009 revealed that the event that recorded the highest number of injury ambulance crashes is the one occurring at the intersection, where the ERV collided with a non-ERV entering the intersection from the crossing street. The event that led to the highest number of property damage crashes was the one where the ERV collided with a non-ERV encroaching into the ERV’s lane from the adjacent one [5]. This research shows that although ERV crashes are notably occurring at the intersection, focusing on the facilitation of their passage should not be limited to intersections but their movement along roads as well.

On the other hand, the efficiency of the ERV operations is directly linked to their response time. When an ERV fails to respond in time to an emergency, the service is deemed inefficient due to the negative impacts that will result. As stated in the Emergency Vehicle Safety Initiative FA-336/February 2014 [6], “adding 10 seconds to a response time is better than not arriving at all because of a collision caused by indiscretion and/or foolishness” p. 86. Excessive delays can also be devastating resulting in unrecoverable impacts. Thus, depending on the type of services they provide, ERVs have specific target response times that should not be exceeded.

Regarding EMS activity, a comparative study about the effectiveness of their response times revealed that the mortality risk was 1.58% for patients provided service in more than 5 minutes compared to 0.51% for those provided service in less than 5 minutes [7]. Also, Pesek [8] discussed the impact that response time has on serious medical emergencies. He showed that when paramedic-treatment starts within 3 minutes of a cardiac arrest, the survival rate stands at 80% but drops to 30-40% when the treatment begins within 8 minutes. Similarly, the American Heart Association (AHA) states that heart attack victims’ chances of survival will decrease by 7 to 10 percent for every elapsed minute [9]. Regarding the fire suppression activity, according to the National Fire Protection Association (NFPA) [10], a response after the flashover (occurring between 4 and 11 minutes after the start of the fire), will most of the times be unsuccessful.

Additionally, based on the Emergency Vehicle Safety Initiative FA-336/February 2014 report [6] and the standards set by the NFPA and the emergency medical services agencies, the target response time of fire and rescue departments and basic life support is 3 to 4 minutes. The target is 8 minutes for at least 90% of the advanced life support requests. Regarding police activity, the Department of Justice sets a 4-minute target response time. As shown in Table 2-1, the ERV response times thresholds are not currently satisfied, and immediate action is tremendously needed.
### Table 2-1: Current response times vs target response times

<table>
<thead>
<tr>
<th>Emergency services</th>
<th>Response time from actual data</th>
<th>Target response time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emergency medical services</td>
<td>The year 2012: 12.57 min (Rural areas) and 7.31 min (Urban areas) [2, 11]  The year 2013: 12.5 min (Rural areas) and 7.14 min (Urban areas) [2, 11]</td>
<td>5 min [7] [8]</td>
</tr>
<tr>
<td>Fire</td>
<td>51% less than 5 min (for fire confined in rooms and floors) 49% less than 6 min (for fire confined in buildings) 54% less than 6 min (for fire beyond buildings) [12]</td>
<td>3-4 min [10] [6]</td>
</tr>
</tbody>
</table>

Given the crash risk and delays, the ERVs’ activities on the roads are not safe nor timely. Hence more attention should be granted to improve their operations, and new methods for improving ERV response should be developed. One method that has been deployed in the field and that still receives research attention is traffic signal preemption and will be discussed in the following section.

#### 2.1.2) Currently deployed measures

The development of methods to enhance the activity of ERVs by reducing the number of crashes and response delays should be considered. Lights and sirens are visual and audible warnings that have long been in use routinely by first responders when responding to and from (in some cases, particularly for ambulances) an emergency scene. Although these warning devices are considered vital, it is also widely acknowledged that their benefits are extremely limited to short-range and slow ERV speeds [14]. On the other hand, the traffic signal preemption is a concept that has been widely studied and elaborated. The latter ensures that an ERV is granted priority over the other vehicle groups when proceeding through a signalized intersection. The traffic signal preemption concept started to be deployed at signalized intersections in the late 1960s [15], after the development of technologies with the ability to detect an approaching ERV. In this section, an overview of the ERV preemption is discussed along with the recovery methods that are needed after each preemption to revert to normal conditions. Benefits to the ERV operations regarding response time and safety has been noted [15] and will be discussed in this section followed by the negative impacts that this technique has on the other vehicles, due to timing plan interruption and insertion of traffic signal plan transition phases.

- **Overview**

  The concept of preemption at signalized intersections is to facilitate the passage of an ERV by giving it priority over other vehicles. The ERV can request an extended green or an earlier green indication to cross the intersection without stopping and with the highest practical speed [10]. According to the
Emergency Vehicle Safety Initiative FA-336/February 2014 report [6], preemption can be activated only at equipped traffic lights after sending a signal from a device installed on the ERV or by remote control at a specific location, such as 911 call centers. Preemption can also be triggered using other ERV technologies such as optical detection and pavement loops [16]. Louisell [10] mentions that the selector installed on traffic signals will scan all the requests before granting the earlier or extended green phase to authorized vehicles. He adds that the preemption signal phases follow the Manual on Uniform Traffic Control Devices (MUTCD) standards and local practices.

- Recovery

Since preemption interrupt the normal cycle and sequence at signalized intersections, recovery or transition phase after each preemption is needed. Transition, by definition, is the practice of switching from one plan to another. Although it is indicated in the MUTC that three phases (“transitioning into,” “serving” and “transitioning out”) should follow when a traffic signal is preempted, there are no established standards and directions regarding which transition strategies to adopt. Transition methods are mostly needed for coordinated traffic signals, where the progression of flow should be highly considered [17].

Several transition strategies exist and differ from one controller to another. The most common transition methods (i.e., available on contemporary controllers) are integrated into microscopic traffic simulation software [17-20]. First, there is the “dwell” or “hold” method in which the green phase is held until it coordinates with the new plan. Then, there is the “max dwell” method in which the green phase will be held but up to a certain maximum duration, spreading recovery over multiple cycles. The “add” or “long” is another common transition method that consists of lengthening the cycle by a maximum percentage of the cycle length. Conversely, the “subtract,” or “short” method consists of shortening the cycle by a maximum percentage of the cycle length. Finally, the “minimax” family (“smooth,” “best way” or “short way”) selects and implements the best transition between the “add” and “subtract” transitions. The transition has been addressed by many researchers previously since it is an indispensable step that should be implemented (1) when dealing with preemption and (2) when switching plans with the varying traffic conditions during the day. Evaluating the different transition methods that are embedded in the current traffic controllers helps quantify the impacts created after the start of the transition phase. Obenberger and Collura [17] emphasized the fact that the impacts of transition strategies should be evaluated to select the most convenient method for every scenario. Common methodologies used for the evaluation of transition strategies are the hardware-in-the-loop simulation (HILS) and software-in-the-loop simulation (SILS). The HILS consists of combining actual traffic signal controllers with a microscopic traffic simulation model while in a SILS, the behavior of traffic controllers is replicated and combined with the microscopic traffic simulation model, eliminating the need for actual controllers and making the evaluation more practical [17]. Some studies focused on transitions methods that specifically exit from
ERV preemption control whereas other studies targeted transition methods in general without specifying the reason behind its need.

Nelson and Bullock [13] evaluated the impacts of ERV preemption because it was difficult to confirm the benefits of preemption due to a lack of quantitative measurements. Using a HILS with standard TSIS/CORSIM simulation, they concluded that the impact of preemption on the overall travel time and delays increase with the number of preemptions. Subsequently, Obenberger and Collura [20] adopted the SILS approach linked to CORSIM to evaluate transition strategies on three levels of traffic volume. They mentioned that the selection of the strategy should highly depend on the current traffic volumes.

On the other hand, Yun et al. [21] favored the comparison of the performance of currently deployed traffic controllers to obtain a more realistic evaluation. Thus, using a HILS that consists of actual traffic controllers (McCain’s 170E and Econolite’s 2070 ATC) and the VISSIM microscopic simulation model, different transition strategies under varying traffic volumes were considered. The fixed exit phase control feature available in the 2070 ATC traffic controllers was also analyzed. This feature allows the selection of specific exit phases to start at the end of the preemption. Yun et al. found that the exit phase feature available in 2070 controllers should be adopted since it showed significant benefits over the 170E controller in which the feature is unavailable. Yun et al. [22] re-evaluated the fixed exit phase control and introduced a dynamic exit phase control available in more advanced traffic controllers. In the dynamic feature, the selection of the exit phases depend on the time of the ERV preemption within the cycle. Finally, Yun et al. [23] compared the performance of ERV preemption transition methods available in 170E controllers (most widely deployed in Virginia). Their results were in line with the ones in [13, 20, 21] where the “smooth” transition methods were the most effective method regarding delays at all traffic conditions and for different traffic controllers. Consequently, several considerations are needed when selecting the transition method as the best universal method (i.e., for all scenarios) does not exist. A review of studies focusing on transitions methods evaluation is presented in Table 2-2 along with the key findings.
Table 2-2: Review of studies evaluating transition methods after ERV preemption

<table>
<thead>
<tr>
<th>Study</th>
<th>Transition method tested</th>
<th>Experimental method</th>
<th>Key findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nelson and Bullock [13]</td>
<td>“Smooth” “Add” “Dwell”</td>
<td>HILS using actual equipment and standard TSIS/CORSIM simulation</td>
<td>Preemption impact on the overall travel time and delays increase with the number of preemptions “Smooth” transition was the most effective “Dwell” transition reflected significant delays</td>
</tr>
<tr>
<td>Obenberger and Collura [20]</td>
<td>“Smooth” (“Best way”) “Add” (“Long”) “Subtract” (“Short”) “Dwell” (“Hold”)</td>
<td>SILS linked to CORSIM</td>
<td>Selection of the strategy depends on the current traffic volumes Outstanding performance of the “smooth” method at all traffic conditions as opposed to the “subtract” method</td>
</tr>
<tr>
<td>Yun et al. [21]</td>
<td>For 170E: “Smooth” (“Shortway”) “Dwell” For 2070 ATC: “Smooth” “Add” “Dwell” Exit phase control in 2070 ATC traffic controllers</td>
<td>HILS using McCain’s 170E and Econolite’s 2070 ATC traffic controllers and VISSIM</td>
<td>“Smooth” transition method was the most effective for both traffic controllers Exit phase feature available in 2070 controllers showed significant benefits over the 170E controllers</td>
</tr>
<tr>
<td>Yun et al. [22]</td>
<td>Fixed exit phase control Dynamic exit phase control</td>
<td>HILS using ASC/3-2100 traffic controller and VISSIM</td>
<td>Fixed exit phase control minimizes disruptions while dynamic exit phase control leads to a further reduction</td>
</tr>
<tr>
<td>Yun et al. [23]</td>
<td>“Smooth” (“Shortway”) “Dwell” (both with different numbers of cycles)</td>
<td>HILS with 170E controllers and VISSIM</td>
<td>“Smooth” with two or three cycles outperforms the “dwell” method (the latter did not achieve coordination)</td>
</tr>
</tbody>
</table>

On the other hand, several researchers evaluated transition methods that are not specifically implemented after preemption (Table 2-3). For instance, transition methods are also used to switch between different time-of-day plans and day-to-day plans to accommodate changes in traffic volumes and ensure adaptive control. Shelby et al. [18] evaluated different transition methods embedded in CORSIM Version 5.2 to determine for each congestion level, the method leading to the best performance regarding average delay per vehicle. However, Cohen et al. [19] claimed that most of the previous studies have evaluated the performance of transition methods based on the absolute delay and travel time and that this does not reflect the actual impacts that transition induces. Thus, they introduced the transient profile analysis method that indicates the distribution of delay over the transition phase.
### Table 2-3: Review of studies evaluating transition methods in general

<table>
<thead>
<tr>
<th>Study</th>
<th>Transition Setting (Preemption, Time-of-day, Day-to-day)</th>
<th>Transition method tested</th>
<th>Experimental method</th>
<th>Key findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shelby et al. [18]</td>
<td>Not specified Plan1 to Plan2</td>
<td>“Dwell” “Max dwell” “Add” “Subtract” “Smooth” (“Minimax”)</td>
<td>Transition methods embedded in CORSIM Version 5.2 simulation</td>
<td>“Smooth” transition effective at all levels of congestion “Add” transition showed some insignificant benefits over smooth methods only under congested levels “Subtract” resulted in minor benefits at low saturation levels “Dwell” and “max dwell” transition led to high delays</td>
</tr>
<tr>
<td>Cohen et al. [19]</td>
<td>Transition between traffic signal plans</td>
<td>“Dwell” “Max dwell” “Add” “Subtract” “Smooth” (“Shortway”)</td>
<td>Transient profile analysis method using CORSIM Version 6.0</td>
<td>“Smooth” transition remained the most effective method. “Dwell” transition showed significant delays again “Add” method turned out to be inapplicable</td>
</tr>
<tr>
<td>Yun et al. [24]</td>
<td>Transition between TOD plans</td>
<td>For 170E: “Smooth” (“Shortway”) “Dwell” For 2070 ATC and ASC/3-2100: “Smooth” “Add” “Dwell”</td>
<td>HILS environment using VISSIM and McCain 170E, the Econolite 2070 ATC, and the ASC/3-2100</td>
<td>“Smooth” transition strategy is the most effective for each traffic controller The 170E traffic controller outperformed the ASC/3-2100 which in turn was more effective than the 2070 ATC due to its faster start in transitioning</td>
</tr>
<tr>
<td>Lee et al. [25]</td>
<td>Not specified (TOD plans, traffic responsive plan selection, ...)</td>
<td>------</td>
<td>Non-linear mathematical model solved with a genetic algorithm</td>
<td>Reduction in delays on side street delays compared to the delays obtained after selecting the transition methods based on simulation</td>
</tr>
</tbody>
</table>

Regarding the smooth and dwell transition methods, the results in [18] and [19] were consistent where the smooth transition is the most effective at all the saturation levels while the dwell transition showed long delays over all ranges of congestion levels along with significant shock waves. Nevertheless, the results regarding the “add” method differed. The results in [19] showed inapplicability of this method while Shelby et al. [18], based on the absolute average delay per vehicle, detected insignificant outperformance of the “smooth” method by the “add” method at high congested levels. Thus, the transient profile analysis method proved to be suitable when examining the impact of transition methods. On the other hand, Yun et al. [24] focused on the performance of vendor-specific transition methods for actual...
traffic controllers (McCain 170E, the Econolite 2070 ATC, and the ASC/3-2100) using a HILS environment and VISSIM. The transition between the time of day plans was examined to determine the controller with the best transition performance. The smooth transition strategy turned out to be again the most effective for each traffic controller. Also, the 170E traffic controller outperformed the ASC/3-2100 which, in turn, was more effective than the 2070 ATC due to its faster start in transitioning. Finally, and unlike all the previously discussed studies, Lee et al. [25] developed a non-linear mathematical model solved with a genetic algorithm and based on optimizing the transition by minimizing delay. This study differs from all previous ones since it does not select the transition method based on simulation. Their approached favored most of the times the major street approaches by giving them more green times.

To summarize, the majority of the studies have been able to model transition strategies using hardware in the loop simulation or software in the loop simulation associated with microscopic traffic simulation software such as CORSIM and VISSIM. Different procedures have been developed to evaluate and compare various transition strategies embedded in contemporary controllers. Some publications only focused on preemption to exit emergency preemption [13, 17, 20-23] while other researchers discussed transition methods on a more general level [18, 19, 24, 25]. Nevertheless, all studies can be considered beneficial since they proved how different transition methods would lead to different levels of severity regarding delays and travel times. Thus, making the selection of the transition strategies is a highly delicate task.

- Evaluation of Emergency Vehicles’ Preemption

The evaluation of the emergency vehicle’s preemption is essential since it ensures that its implementation will benefit the overall traffic operations. A study conducted by Virginia’s Department of Transportation in 1997 [26] showed that, at that time, only 5.27% of the agencies with deployed signal preemption have used that feature, and the reasons behind this included: lack of need, minor benefits for emergency vehicles, impacts on other vehicles and poor understanding of preemption benefits. Louisell et al. [27] clarified that the problem with preemption implementation is the lack of standardized quantitative evaluation of its performance (i.e., resulting in benefits to ERVs operations and impacts on other vehicles). Such standardized assessments are needed to encourage the adoption of preemption at traffic signals. As a result, several models and evaluations were developed afterward to evaluate the benefits and impacts that preemption for emergency vehicles can produce [10, 26-29].

- Benefits for Emergency Vehicles

Preemption is a concept that has been mainly developed to improve the ERV’s activity. It aims to provide an undelayed passage of ERVs at intersections at maximum practical speed, with little risk of vehicle collisions. Subsequently, reductions in crashes and travel time delays are certainly expected. However, the amount of benefits is not the same under all conditions and differ from one situation to another.
depending on several factors such as the number of preemptions in the network, the traffic volume levels, and the ERV type. Louisell et al. [27] affirm that preemption benefit ERVs operations regarding reduced crash potential and reduced travel times. Subsequently, Louisell [10] developed a technique that associates crash potential with conflict points at signalized intersections so that safety benefits estimates can be obtained.

Regarding the travel time benefits, the results also indicated that preemption allowed an ERV to travel a single intersection located on an arterial in Northern Virginia with an average time savings of 30 seconds. Additionally, another assessment at a single intersection with one preemption request showed that the time savings would increase when moving from low to medium volume levels [26]. Finally, the Emergency Vehicle Safety Initiative FA-336/February 2014 report [6] highlighted that preemption at an intersection result in a more significant reduction in response time for large vehicles versus police cars since the latter can stop and accelerate in a shorter interval of time.

- Impacts on other vehicles

On the other hand, ERV’s preemption impacts the operations of other vehicles. For instance, vehicles coming from the conflicting approaches are requested to stop to ensure the safe passage of the ERVs. This interruption in the regular traffic cycle creates delays on the conflicting approaches. The level of severity depends on several factors such as intersection spacing, transition algorithm, traffic volumes, duration of preemption and available slack time during cycles [10]. According to Nelson [13], a study at four signalized intersections along the main arterial proved that a single preemption request has minor effects on the travel times of other vehicles. The same study showed that multiple preemptions at closely spaced time intervals would lead to significant delays that can reach 20 to 30 seconds per vehicle. Though the results were site-specific, similar evaluations can be replicated for any other location. In McHale [26], the time impact to non-ERVs ranged between 1.1% and 3.3% increase in their overall travel time.

Nevertheless, this increase is relatively minor if compared with the savings in ERV response times. Similarly, Gkritza [3] claimed that the impact on the other vehicles is low since the disruption caused by preemption is short. The average duration of preemption ranges between 16 and 26 seconds and, most of the time, an emergency response only includes one ERV (73% of the cases in DC).

To summarize, preemption aims to facilitate the passage of the ERVs at intersections. It consists of requesting an extended green time or an earlier one to be able to cross the intersection without stopping and with the highest speed possible. However, when preemption is implemented, one should not only consider the benefits to the ERV operations regarding reduced travel times and delays. Preemption impact the operations of the non-ERVs. Although the impacts are not significant, researchers are always interested in conducting studies to evaluate preemption and to optimize its implementation under different conditions. When preemption is considered, the passage of the ERVs at intersections is facilitated. However, the ERV
movement from origin to destination should be considered, and this involves the enhancement of the passage along arterials as well. The latter cannot be provided using the emergency preemption technique. Therefore, new techniques should be developed to optimize the operations of the ERV along its complete trip.

2.2. Overview of the Connected Vehicle Environment

With the rapid advancement in wireless and communication technologies, surface transportation has dramatically evolved. With the aim of providing higher mobility, enhanced safety, and less environmental impacts, Intelligent Transportation Systems (ITS) program has been established, and modern transportation started growing. As a part of the ITS program, the Connected Vehicles (CV) and the Autonomous Vehicles (AV) environments emerged. The autonomous vehicle technology mainly uses advanced positioning systems, such as GPS and LIDAR, in addition to built-in sensors to detect and identify its surroundings. Subsequently, decisions will be generated to move and maneuver the vehicle without any human intervention and control along arterials and across intersections. The connected vehicle environment is different from the AV environment since it is a concept based on the utilization of advanced information and communication techniques so that vehicles, infrastructure, and pedestrians can cooperate and communicate. According to Goel et al. [30], the connected vehicle technology will enable the exchange of information and data by allowing vehicles to communicate with each other using vehicle-to-vehicle communication (V2V) and with the roadway using vehicle-to-infrastructure communication (V2I). The main factor behind the wide variety of benefits brought by this new environment is the short latency and accuracy of the real-time information that can be broadcasted through V2V and V2I communications. This information includes real-time data such as location, speed, direction, braking status of all nearby vehicles, and infrastructure information.

The connected vehicle environment primarily relies on dedicated short-range communication (DSRC). The latter is considered suitable for V2V and V2I in the connected vehicle environment since it is characterized by fast network acquisition, low latency, high reliability, priority for safety applications, interoperability and security and privacy [31]. Furthermore, the new generations of cellular networks (5G and LTE) that are already widely deployed can also complement the role of DSRC and even overcome the DSRC challenges mainly associated with the limited communication range and availability [32]. A wide variety of algorithms and systems, such as situational awareness, crash avoidance, and driving assistance systems, has been developed to provide safe and optimal performance on the roads. Improving safety and mobility is not the only gain. For instance, this new environment will also increase the efficiency in traffic operations and will offer a wide variety of environment-friendly services ([30, 31, 33]). Connected vehicle technologies are expected to lead to the optimization of fuel and significant reductions in pollutant
emissions. For instance, new systems such as adaptive traffic signal control at intersections will be implemented and will result in a wide range of environmental benefits.

Additionally, an increase in the overall efficiency is expected due to the reduction in the number of stops and the delays encountered by vehicles ([34], [35]). Aparicio et al. [35] added that this technology would be able to assist first responders. They explain that it will increase first responders’ efficiency and reduce the number of accidents in which emergency vehicles are involved.

There are challenges that need to be addressed to be able to fully leverage the services that the connected vehicles can offer, some. Chan [31] asserted that challenges such as market penetration, security, durability, life consistency, and driver distraction exist. Also, Chan stated that implementing this concept has already presented a dilemma since the deployment of the DSRC will take too much time and will demand considerable investments. Aparicio et al. [35] added that this new platform would have to overcome challenges related to compliance with standards, immediate intersection recognition, multiple vehicle priority classifications as well as legal issues that differ from one location to another [36]. Fang [33] and Jimenez [37] asserted that the massive amount of information that will be broadcasted and exchanged through V2V and V2I communications could produce technical as well as practical problems. So, Fang continued and claimed that this would require an effective data management system to ensure the feasibility and sustainability of the overall system. Finally, this environment is subject to multiple security threats, and these could undoubtedly affect its availability, authenticity, and confidentiality [33].

2. 2. 1) Applications at intersections

Intersections are considered major conflict points and act as bottlenecks resulting in significant delays. Therefore, it is essential to leverage connected vehicle technologies to enhance traffic operations at signalized and unsignalized intersections. Several studies introduced models that can be implemented at intersections while utilizing the V2V and V2I communications. In this section, applications for signalized and un-signalized intersections are presented.

It is widely acknowledged that fixed or pre-timed traffic signal plans are not efficient. They rely on expected demands on approaches and repeat the same cycle with the same phase sequence and splits. Thus, they do not respond to any variations in volumes [34, 38]. As a result, actuated signal controls were introduced to optimize the traffic flow and adapt the traffic signal plan to current volumes using vehicle detection methods. It is true that point sensors can detect approaching vehicles. However, these traditional devices can only provide limited vehicle data at fixed locations making it difficult to optimize the actuated traffic signal plan [39] accurately. Improved vehicle detection methods are therefore required.

With the rise of the connected vehicle environment, several studies (shown in Table 2- 4) focused on optimizing the traffic controls at signalized intersections by developing adaptive traffic signal controls that rely on real-time information collected using V2V and V2I
communications [34, 38-41]. Priemer et al. [39] developed an approach that consists of determining the optimal phasing sequence that reduces the predicted total queue length using dynamic programming and complete enumeration. Similarly, the information from connected vehicles was used in the Cumulative Travel-Time Responsive (CTR) real-time control algorithm developed by Lee et al. [40]. It consists of determining the cumulative travel time (CTT) of the vehicles waiting at each phase of the intersection and granting an earlier green phase or an extended one to the phase with the highest CTT. Priemer et al. [39] and Lee et al. [40] acknowledge that assuming 100% market penetration of connected vehicles will not reflect real-life conditions in the short future. Therefore, Priemer et al. [39] used a ql-estimation technique, and Lee et al. [40] used the Kalman filtering technique to account for unequipped vehicles crossing the intersection. Another approach targeting signalized intersections is called the Predictive Microscopic Simulation Algorithm (PMSA) and is introduced by Goodall et al. [38]. It only uses V2V and V2I communications to collect a snapshot of positions, speeds and heading of the equipped vehicles. This information was used to minimize an objective function that reflects delay or a combination of delays, a number of stops and other efficiency parameters to optimize the phase sequence at the intersection. In He [41], a Platoon-based Arterial Multi-modal Signal Control with Online Data (PAMSCOD) was developed. This study optimized the performance of multiple travel modes at an intersection using V2V and V2I communications and also enabled the identification of platoons. Information was collected from the green light requests received from the approaching vehicles, and a mixed-integer linear program was executed to obtain an optimal signal plan for the existing traffic. Gradinescu et al. [34] differed from the previously discussed studies in which the adaptive traffic signal control was decentralized at each intersection. They leveraged the exchange of information between traffic controllers. Retrieved data from equipped vehicles was collected and then sent from one intersection to another along an arterial so that the downstream controller determines its signal plan (cycle length and green times for each phase) for the following cycle based on the actual volumes. These studies yielded promising improvements regarding delays, stops, speeds and queue lengths when compared to pre-timed signals and to actuated signals that are not relying on accurate real-time information.

Table 2-4: Adaptive traffic signal control using connected vehicles

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Objective</th>
<th>I2I?</th>
<th>Data collection</th>
<th>Methodology</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gradinescu et al.</td>
<td>2007</td>
<td>Adaptive traffic signal plan</td>
<td>optimize the cycle length and phase at the downstream controller once every cycle by minimizing delay</td>
<td>exchange of information between controllers</td>
<td>V2V and V2I, VANET discrete-event simulation tool</td>
<td>Reductions in delays &amp; queue lengths along with significant operational and environmental benefits when</td>
</tr>
<tr>
<td>Authors</td>
<td>Year</td>
<td>Plan Type</td>
<td>Optimization Approach</td>
<td>Simulation Tools</td>
<td>Market Penetration Requirement</td>
<td>Results</td>
</tr>
<tr>
<td>------------------</td>
<td>------</td>
<td>--------------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------</td>
<td>---------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Priemer et al.</td>
<td>2009</td>
<td>Adaptive traffic signal plan</td>
<td>Optimize the phasing sequence at an intersection by minimizing total queue length</td>
<td>Loop detectors and V2I</td>
<td></td>
<td>Improvements reflected by a total delay reduction up to 24% and a mean speed increase of 5% if compared to pre-timed and actuated signals</td>
</tr>
<tr>
<td>Goodall et al.</td>
<td>2013</td>
<td>Adaptive traffic signal plan</td>
<td>Optimize the phase order by minimizing an objective function that reflects delay or a combination of delays, number of stops and deceleration</td>
<td>Decentralized V2V and V2I</td>
<td></td>
<td>When compared to an actuated timing plan, improvements in delay and speeds during unexpected demands</td>
</tr>
<tr>
<td>Lee et al.</td>
<td>2013</td>
<td>Adaptive traffic signal plan</td>
<td>Cumulative Travel-time Responsive (CTR) real-time control algorithm that optimizes the phase sequence by giving priority to the combination of movements with the highest cumulative travel time</td>
<td>Decentralized V2V and V2I</td>
<td></td>
<td>Improvement in the total delay and average speed at the intersection when at least 30% market penetration is available, compared to an optimized actuated timing plan</td>
</tr>
<tr>
<td>He et al.</td>
<td>2012</td>
<td>Adaptive traffic signal plan</td>
<td>Platoon-based arterial multi-modal signal control that optimizes the phase sequence by minimizing an objective delay function</td>
<td>Decentralized V2V and V2I</td>
<td></td>
<td>Improvement in delays for vehicles and buses under non-saturated and oversaturated traffic conditions compared to an actuated signal control</td>
</tr>
</tbody>
</table>

On the other hand, vehicles which are in the dilemma zone (i.e., confused with the stop/go decision making when approaching a signalized intersection) are sources of disturbance and subject to collisions. In Zha et al. [42], an integrated dilemma zone protection system that exploits the connected vehicle environment has been introduced. The system uses V2I communications and defines the green time extension that should be granted to vehicles in the dilemma zone after taking into consideration the resulting impacts on the vehicles on the conflicting paths. This system (at full market penetration), when compared to the traditional green time extension systems, could protect the vehicles in the dilemma zones more, while
also reducing the delays encountered by the conflicting movements. Due to V2I communications, the approaching equipped vehicles can receive a message indicating when the light will be turning green.

Other studies focused on improving the traffic operations at intersections using the connected vehicle technologies by optimizing the vehicles’ trajectory. These studies did not consider the presence of traffic signal control; thus, they indirectly address the case of unsignalized intersections. For instance, an intersection is considered unsignalized if no control or traditional control systems (such as the traffic signal control and yield signs) exist. Achieving safe and efficient traffic operations at these intersections is challenging due to confusion, human errors, frequent stops and unnecessary idling [43]. Thus, several studies addressed the operations directly at unsignalized intersections by using wireless communication between vehicles without relying on the presence of traffic signal lights. First, Li [44] introduced a system that provides collision-free movements at blind crossings after the enumeration of all safety driving patterns in a spanning tree and the elimination of all infeasible ones. Trajectory planning for the feasible patterns was then determined and the trajectory that requires the least overall travel time was adopted.

Similarly, Guler et al. [45] discussed a model that focuses on optimizing the total delay at an isolated intersection by listing all possible discharge sequences and selecting the optimal one minimizing an objective function for total delay or the total number of stops. This approach also utilizes information from the connected vehicles but also relies on platoon discharging. Besides, Lee et al. [43] presented the Cooperative Vehicle Intersection Control Algorithm (CVIC) that enables cooperation between vehicles and infrastructure for an effective activity at unsignalized intersections as well as signalized intersections. It was developed so that the overlapping trajectories of vehicles crossing an intersection is minimized. Mourad et al. [46] developed a model based on the Timed Petri Nets with Multipliers (TPNM) that minimizes the total queue lengths at an intersection and displays a green or red signal to indicate whether the vehicle can proceed through the intersection or not. They assume that all vehicles are equipped with onboard units that broadcast information such as position and speed. The Cooperative Adaptive Cruise Control system is an advanced cruise control system that relies on V2V and V2I communications. By developing this system, researchers primarily targeted mobility on highways. Zohdy et al. [47] adapted the approach to intersections. They developed a game theory-based algorithm that consists of predicting the trajectories of the vehicles entering an intersection and controlling their speed in a way to avoid conflicts and to minimize the total delay.

The promising safety and mobility benefits that these connected vehicles applications implemented at intersections are capable of offering to all vehicle types are the reason behind the interest that researchers have in developing models that tackle the challenges experienced at intersections by specific vehicle types such as the transit vehicles.
Transit signal priority (TSP) is a concept that offers priority to transit vehicles when crossing signalized intersections. This concept has been thoroughly examined by researchers and strategies that evaluate the effects of its deployment are developed. According to Koonce [48], TSP can offer considerable benefits regarding travel time reliability and the reduction of delays experienced by transit vehicles. Besides, it can also provide significant savings characterized by a reduction of pavement deterioration, maintenance, and operating costs. However, the implementation of TSP has been limited due to several technical and financial reasons. Hu et al. [49] assert that if it is deployed with inaccurate data and improper models, it might create negative impacts that outweigh the resulting benefits offered to the activity of the transit vehicle and nearby vehicles. Koonce [48] highlights the idea that connected vehicles can be easily merged with the TSP systems and Hu et al. [49] continue that the high accuracy of the real-time information available in the connected vehicle environment could be a potential factor behind the success of the TSP deployment in the future. Hu et al. introduced a new transit signal priority logic that uses accurate information exchanged by V2V and V2I communications. It consists of reallocating the green time instead of adding an interval of time for the passage of the transit vehicle. After evaluating this new system regarding delay per person and transit delay, this new approach showed considerable improvements with no adverse effects when compared to the conventional TSP and the no TSP case.

Similarly, Zeng et al. [50] developed a model called the person-based adaptive signal priority control that is also based on information sent from connected vehicles and consists of minimizing the person delay. Simulation and sensitivity analysis results showed that, when compared with SYNCHRO’s optimized timing plan, a decrease in the overall delay is noticed along with a more significant reduction in the transit vehicles’ delays in scenarios with at least 30% market penetration. To summarize, the TSP is a practice that improves the activity of a particular type of vehicle; yet, it will inevitably engender delays for the vehicles on the conflicting approaches. Therefore, relying on advanced technologies such as V2V and V2I communications can provide real-time and accurate information that results in an improved TSP deployment that focuses on alleviating the impacts experienced by the other nearby vehicles.

As shown in the discussion above, researchers are interested in improving the traffic operations at signalized and unsignalized intersections as well as enhancing the TSP systems. Various studies have shown that connected vehicle technologies could be integrated to improve the performance at intersections. The benefits offered by the connected vehicles are not limited to the improvement of traffic operations at intersections. The integration of connected vehicle technologies in models focusing on the enhancement of traffic operations along transportation links will be discussed in the following section.

2.2.2 Applications at links

Driving along arterials is a combination of complex tasks involving the adaptation of the vehicle’s speed to different factors such as surroundings, lane changing, ramp merging. This complexity is a reason
behind the manifest delays and a high number of crashes in our transportation system. Several studies utilized connected vehicle technologies to enhance mobility as well as safety along arterials to alleviate the challenges. First of all, the connected cruise control systems are introduced by Jin et al. [51] and Orosz [52] with the aim of assisting equipped vehicles to adapt to the varying traffic conditions after receiving traffic information from downstream vehicles via V2V communications. Results showed that connected cruise control succeeds in guaranteeing a uniform flow and achieving enhanced string stability (i.e., reduction in velocity fluctuations) as going from the head to the tail of vehicle platoons. Another mobility model was discussed in Tian et al. [53] in which vehicles move cooperatively as a group while keeping a minimum distance between them to avoid a collision. For instance, the model introduces the cooperative behavior of connected vehicles inspired by the fish school concept. In other words, attractive forces keeping the vehicles in groups, as well as repulsive forces taking into account constraints such as road obstacles and boundaries, are applied on the vehicle to direct it along the arterial. The model strongly relies on the use of V2V communications and achieved considerable benefits regarding safety as well as the efficiency of vehicles’ operations. Moreover, the performance of vehicles has also been targeted by Yu et al. [54] who developed a car-following model that relies on the exchange of information (e.g., headway, relative speeds) among vehicles to reduce traffic jams and offer enhanced traffic stability.

The availability of real-time information in a connected vehicle environment can also improve the mobility of vehicles by assisting them when selecting the lane to be used to ensure a safe and smooth movement. A lane changing movement is directly linked to the lane selection decision-making which can potentially deteriorate the traffic operations when no external guidance is provided. Therefore, optimizing the lane selection using connected vehicle technologies is of great interest. For example, Jin et al. [55] proposed a real-time optimal lane selection system that consists of three steps: a data collection of information about the vehicles within the V2V and V2I communication range using a lane selection agent, a determination of the optimal target lane and a lane changing implementation. The selection of the optimal lane is based on a model whose objective function minimizes the overall number of conflicting jobs. The approach, simulated using SUMO, resulted in reductions regarding travel time, energy consumption and emissions of pollutants, proving that the exchange of real-time information between vehicles can offer situational awareness that will help vehicles make effective operational decisions.

V2I wireless communication can also benefit a vehicle merging into the traffic from a ramp lane. This difficult task often results in a collision due to confusion and is facilitated by an automated on-ramp merging control algorithm [56]. Based on an exchange of information (e.g., ID, location, speed, and intent between nearby vehicles), the model ensures a collision-free on-ramp merging by notifying the merging vehicle when there is enough space between vehicles and by generating the corresponding required acceleration/deceleration actions to adopt by this vehicle.
Mobility on highways has always been an issue characterized by traffic jams in major cities and especially during peak hours. Therefore, researchers were interested in improving the mobility of vehicles on arterials using the capabilities of the connected vehicle environment. Acceleration-based mobility models, lane changing and ramp merging models leveraging the connected vehicle technologies are the main applications targeting traffic operations on transportation links. In the following section, other connected vehicles applications that go beyond traffic operations improvements are discussed.

2.2.3) Other applications: accurate measurements

One useful capability offered by the connected vehicle environment is the ability to provide accurate measurements of data that are not easy to obtain. For instance, the exchange of information among vehicles and between vehicles and infrastructure led to the development of multiple models for crash identification, travel time estimation, and pavement evaluation. First, Talebpour et al. [57] developed a model that identifies the number of near-crashes using trajectory data broadcasted by the connected vehicles by comparing the acceleration of a vehicle to a pre-determined threshold. If the acceleration is lower than the threshold, a potential near-crash is identified and the time to collision is determined. If the latter is lower than the follower’s reaction time, a near-crash is documented. Results met the expectation that the number of near crashes increases with traffic density. Road accidents mostly affect the traffic operations and impact the travel time reliability of the vehicles approaching the accidents as well. The connected vehicle technologies have contributed to obtaining a more accurate travel time reliability by assisting the drivers while selecting the best route. Second, a travel time reliability (TTR) model was developed by Lei et al. [58] to calculate the travel time reliability on a freeway where an accident occurred. The TTR was estimated by gathering, from the connected vehicles, information about the accident such as time delay, position, duration. Similarly, Tian et al. [59] developed a dynamic travel time estimation model that is crucial for route guidance applications since traffic conditions will constantly change with time. The data collected from the connected vehicles accurately reflect the traffic variations which can be computed using a road link dynamic algorithm.

The connected vehicle environment can also be highly beneficial to transportation asset management practices by providing accurate information to assess the location and level of deterioration and distress in the pavement. Dennis et al. [60] highlighted how crowdsourced data from connected vehicles could be utilized in a framework for the evaluation of pavement conditions. Also, Bridgelall [61] introduced the road impact factor that can be computed from these collected data and he showed that this factor is directly proportional to the international roughness index which is more difficult to compute. In Bridgelall [62, 63] theoretical precision boundaries for the road impact factor were identified based on several case studies.
Consequently, the connected vehicle environment has benefits from improving the traffic operations on intersections and arterials to obtaining accurate data that is not easy to measure. This new environment has already been widely integrated into various areas with the aim of drastically improving transportation systems. The following section introduces several connected vehicles applications that potentially improve the emergency response vehicles’ operations.

2. 3. Connected vehicles applications assisting the ERVs movement

As previously discussed, extensive research has been already conducted to develop models and applications that improve the overall traffic operations and as a result of the daily road experience by relying on the connected vehicle technologies. This emerging environment has already been proven beneficial to transportation systems on several levels. Regarding the ERV’s activity, its safety and efficiency had always been questioned. The characteristics of the current ERV operations (discussed in Section 2. 1. 1) demonstrate why the currently deployed practices that assist the ERV while traveling along roads are insufficient. Thus, there is a great interest in benefitting from the future connected vehicles technologies to improve the ERV’s movement on the roads. For instance, researchers have already targeted the movement of ERVs explicitly or implicitly, with the aim of making it safer, smoother and faster-using models based on the new wireless and communication capabilities of the connected vehicle environment.

2. 3. 1) At intersections

As previously discussed, the movement of ERVs at intersections is the most critical as it is characterized by the most significant delays and crashes. The ERV preemption practice, discussed in Section 2. 1. 2), is widely deployed nowadays and consists of granting a green signal to the ERV approaches at signalized intersections. With the improvements in wireless and communication technologies, researchers are improving the ERV detection techniques. For example, two systems are proposed in [64] [65] and both rely on the use of the radio frequency identification (RFID) technology (i.e., a system that consists of tags, readers and embedded applications) to identify an approaching ERV and grant it a green light in a timely manner, depending on the level of congestion. Another model that uses V2V communications and that ensures a faster ERV passage at congested signalized intersections has been proposed [66, 67]. It consists of instructing the vehicles on the ERV’s approach to following a specific behavior to clear a lane for the ERV allowing it to proceed through the intersection at its maximum speed. The vehicles’ behavior is determined based on an approach that splits the vehicle queues at critical points, which are determined based on the kinematic wave theory.

On the other hand, autonomous intersection management is proposed in [68-70] where all vehicles are considered as autonomous agents (equipped with sensors) which enter the intersection by following a reservation-based approach granting priority to ERVs. Additionally, an adaptive traffic intersection system
based on the Wireless Sensor Network is developed in [71]. It consists of gathering information about traffic using sensors installed along the roads then prioritizing the passage of ERVs through communication between traffic lights.

2. 3. 2) At links

For transportation links, Toy et al. [72] proposed an approach that assists ERVs while traveling on automated highways. It consists of forming gaps by moving the platoons of vehicles left, right or forward to provide enough space for the ERV to proceed rapidly through congested areas on highways. Moussa [73] relied on a cellular automata model to facilitate the passage of the ERVs but without explicitly assuming the use of V2V or V2I technologies. The ERV is requested to move to the lower density lane while the downstream vehicles, depending on their location concerning the ERV, may have to change lanes to clear the path. Another approach is developed by Yoo et al. [74] based on a road-reservation approach for ERVs using sensor networks. The non-ERVs received instructions to move away from the dedicated ERV path to promote a safe and fast passage.

A VANET-based emergency vehicle warning system [75] has been designed by mainly relying on inter-vehicle communications and roadside infrastructure. The system ensures that downstream vehicles are well aware of the approaching ERV. For instance, a warning message including all relevant information about the ERV such as location, speed, position, and route will be broadcasted from the ERV to the nearby road users. Roadside equipment is receivers that will indirectly warn the non-equipped vehicles that are close to the EV path. Similarly, a data flow design for emergency warning systems based on VANET network has been proposed [76]. This study provides communication between the ERV and neighboring vehicles can enhance the drivers’ decision-making to reduce the ERV’s response time and enhance its safety. Vehicles receive a warning message that includes detailed information about the ERV route; subsequently, a decision algorithm is executed to determine the appropriate actions. A more advanced warning also computes the optimal EVR route while taking into account the variation in the traffic conditions [77]. This centralized model executes an ERV dynamic path arrangement followed by a data dissemination strategy to warn the downstream vehicles of an approaching EV. Using the DSRC network, roadside units installed at every intersection collect, in a periodic manner, the real-time traffic information. Using an Ethernet network, the information is sent to the centralized server that start the path arrangement. This step consists of implementing the A* algorithm to determine the primary shortest path and another secondary path. The latter is a path that is used in case unexpected events happen on the primary one. Therefore, it should not be overlapping with the primary path and sub-paths connecting the primary to the secondary with minimum time is determined as well. These set of paths are sent back to the roadside unit (RSU) which in turn sends them to the ERV. The centralized server periodically evaluates the shortest path from the ERV location to its destination, notifying it about any beneficial path adjustments to handle the
dynamic aspect of traffic. Once the path of the ERV is set, the centralized server calculates a Controlled Emergency Distance (CED) which delineates the area that can affect and obstruct the ERV’s path. Warning messages are generated and sent to the RSU in the CED which, in turn, send them to nearby vehicles.

2.3.3) Other emergency assistance systems

With the introduction of the new wireless and communications technologies, several aspects of emergency response can be improved. Automatic crash notification [78-80], for example, shortens the time to notify dispatchers of an incident. Geographic Information Systems (GIS) have also been leveraged [81-83], and an emergency rescue system in a freeway environment has been proposed [81] to help with crash location and response vehicle routing. Introduced by the US Department of Transportation (USDOT), the new program for Response, Emergency Staging and Communication, Uniform Management and Evacuation (RESCUME) focus solely on the enhancement of the emergency activity [84]. For instance, staging guidance, advanced automatic crash notification, and emergency responder vehicle dynamic routing are part of this program. Connected vehicles technologies can also be very advantageous to ERVs during evacuation situations. An intelligent disaster management system has been proposed with the aim of enhancing the transportation evacuation strategies [85]. It uses Vehicular Ad hoc Network, V2V, and V2I communications to provide an exchange of relevant information needed to generate evacuation strategies. Another VANET-based approach is developed with the aim of reducing traffic chaos and confusion and enhancing the emergency response during evacuation situations [86]. Information about the traffic is periodically gathered from the vehicles on the roads in addition to information related to the ERV response such as current location, destination, and path.

The search and development of new solutions to improve the operations under emergency response situations are highly needed. The traditional emergency preemption concept aims to facilitate the passage of the ERVs at intersections and consists of requesting an extended green time or an earlier one so that the ERV can cross the intersection at the highest speed possible. Nevertheless, preemption only facilitates the ERVs’ passage at intersections and does not consider their movement on transportation links. Therefore, there is a need for new technologies and new concepts that can improve and facilitate the movement of the ERVs along their complete routes from origin to destination. The connected vehicles environment’s technologies are capable of promoting an increased efficiency at signalized and unsignalized intersections as well as better mobility along transportation links. The broadcast of real-time information and the V2V and V2I communications can play a significant role in achieving a fast and safe emergency response.
Reference

2. NHTSA, TRAFFIC SAFETY FACTS 2013. 2013, NHTSA.
11. NHTSA, TRAFFIC SAFETY FACTS 2012. 2012, NHTSA.


32. 5G Americas, V2X Cellular Solutions. 2016.


Emergency response vehicles’ (ERVs) travel is risky, as non-ERV drivers are often unsure of the ERV’s next maneuver and how to facilitate its movement. An integer linear program (ILP), introduced in this paper, facilitates the ERV’s movement through a transportation link. Leveraging vehicle-to-vehicle communications, information is collected about vehicles on a link section. Then, the ILP finds the ERV’s fastest intra-link path. To increase safety, the ILP assigns non-ERV locations as far away from the ERV as possible while avoiding passing and weaving among vehicles. The ILP can be adapted to different ERV sizes, road types, surrounding conditions, etc. Sensitivity analysis indicated that scenarios with narrower road segments and higher numbers of non-ERVs led to ERV paths with lane changes and higher computation times. When compared to current practice requiring non-ERVs to move to the nearest road edge when an ERV with lights and sirens is noticed, the proposed formulation improved the ERV speed while reducing the conflicts and confusion experienced by downstream vehicles.

3.1 Introduction

Emergency response vehicle (ERV) travel needs to be fast, which is difficult and potentially unsafe in congestion. Risky maneuvers and other drivers’ responses to ERVs has led to thousands of annual crashes [1] and hundreds of line-of-duty deaths [2]. First responders seek to arrive quickly to maximize their intervention’s effectiveness. If firefighters arrive after the flashover (4 - 11 min after the fire starts), the response is likely to be unsuccessful [3]. High response times are most likely in urban areas during periods of heavy congestion. Drivers near the ERV path experience confusion [4] that limits the effectiveness of their efforts to facilitate the ERV’s progress.

ERV movement could be facilitated on links and at intersections. ERV signal preemption has been implemented and continues to evolve. However, opportunities still exist to reduce confusion and ERV travel time on links.

This paper introduces an integer linear program (ILP) that leverages the connected vehicle environment. After collecting information about downstream vehicles (location and speed), this ILP finds the intra-link path that maximizes the ERV’s speed and the free space around the ERV’s path. Maneuvering
instructions are incrementally disseminated to the ERV traveling through the link and to downstream vehicles requesting them to stop at specific positions along the link.

3.2. Literature review

Aspects of emergency response, such as automatic crash notification [5-7], crash location, and routing [8-10], have improved with new technologies. The US Department of Transportation’s RESCUME program enhances emergency activity [11] through staging guidance, advanced automatic crash notification, and ERV dynamic routing. Buchenscheit et al. [4] proposed a system to warn non-ERVs of an approaching ERV, relying on vehicle-to-vehicle (V2V) communications. Rizvi et al. [12] predicted if, when and where vehicles conflict with the ERV’s path to send warnings. Huang et al. [13] developed a warning system using the DSRC protocol. Bratu et al.’s approach displays warning messages using global system for mobile communications modules on electronic boards along the ERV’s path [14]. While these studies emphasized potential benefits of disseminating early warnings to vehicles that might encounter the ERV path, they do not provide the actions that non-ERVs should take.

ERV facilitation efforts can be divided into two categories: optimization (1) at intersections and (2) on links. Intersections are conflict locations. Some studies used communications to improve traffic operations at intersections in general (e.g., [15] and [16]). Others specifically addressed the ERVs’ passage at intersections; reservation- [17], kinematic wave theory- ([18] and [19]), and distance-based [14] approaches have been proposed along with V2X technology use. For multiple intersections, Kamalanathsharma and Hancock [20] introduced a control system that adjusted ERV preemption at downstream intersections based on clearance times and queue lengths. More relevant to this research are studies making the ERV movement on a link safer, smoother and faster. Weinert and Düring [21] introduced a V2V application to provide a free lane, operationalizing the cooperative behavior developed by Düring and Pascheka [22]. Similarly, Toy et al.’s [23] approach moves platoons to provide space for the ERV. In these studies, the ERV path is assumed to be known by downstream vehicles, so the best ERV maneuvers are not dynamically adjusted based on downstream traffic conditions.

Yoo et al. [24] introduced a road reservation approach where non-ERVs received instructions to move away from the ERV path on the lower density lane. Likewise, Moussa’s [25] cellular automata-based approach instructs non-ERVs positioned on a two-lane link to move away from the lower local density lane before pulling over, but without explicitly relying on V2X technologies. In this paper, the pulling-over of downstream non-ERVs is also considered but without assuming that the ERV is traveling on the lower density lane since the ERV may want to turn or stop on a specific lane for an incident. Thus, freeing the lowest density lane might not always be effective. The most closely related systems is introduced by Djahel [26] which discussed only the vision (but no details) of communicating new driving policies (e.g., speed
limit change) to vehicles downstream of an ERV.

The proposed system computes the optimal intra-link ERV path based on downstream conditions. It is not merely a warning system as it also optimizes the locations of downstream non-ERVs by instructing them where to stop. Although this paper only focuses on facilitating the passage of a single ERV along a link segment, it provides a platform for future extensions (e.g., the passage at intersections).

3. 3. Problem statement and assumptions

3. 3. 1) Problem Statement

Given a directed link segment divided into identical cells (length $L$ and width $W$) with $|J|$ non-ERVs traveling along this segment, find the set of cells constituting the ERV’s path which provides the fastest passage while assigning each non-ERV to a specific cell based on its feasible stopping distance.

3. 3. 2) Roadway Discretization and Labeling

This study relies on a specific coding of the cells and vehicles (see Figure 3-1). The $x$-axis represents forward motion while the $y$-axis represents lateral motion (e.g., lane changes). The number of lateral cells ($Y$) depends on the road’s width, including shoulders and other traversable surfaces. Cells are labeled in the ‘$x$’ direction with 1 being closest to the ERV and increasing with longitudinal distance. In the ‘$y$’ direction, cells are labeled in ascending order from the right lane to the left. The non-ERVs are indexed by $j$ which increases with the ‘$x$’ position. If two vehicles are located at the same ‘$x$’ position, they are labeled in ascending order from right to left. The ERV’s speed and instruction variables are superscripted by $i$ since they are given at every increment $i$. An increment is a longitudinal distance (along the $x$-axis) that is equal to the ERV’s longitudinal size ($N$) plus a buffer. Here, the buffer is 1 cell, so the increment encompasses $(N+1)$ longitudinal units.

![Discretization of roadway network](image)

Figure 3-1: Discretization of roadway network

3. 3. 3) System Description

The non-ERVs of interest are located within an initial range (IR), which is at a pre-defined distance from the ERV, as shown in Figure 3-2. This approach consists of collecting information from these non-ERVs and, after a time interval $\Delta t$ (computation time), broadcasting the non-ERV assignment (final)
locations. For negligible $\Delta t$’s, non-ERVs travel a small distance that can be disregarded. Otherwise, their positions and speeds after $\Delta t$ will be estimated since this formulation requires the non-ERV locations $(x'_j, y'_j)$ and speeds ($\sigma_j$) at the time they receive the message instructing them where to stop.

Each non-ERV should be assigned to a cell it can reach based on vehicle dynamics [27]. This observation forms the basis of how the feasible stopping range (FSR) for each non-ERV $j$ is defined. The FSR of each vehicle $j$ starts at a minimum final position ($MFP_j$) which is identified, using (1) and (2), based on $j$’s initial longitudinal index ($x'_j$) and its minimum stopping distance ($MSD_j$), in cells, that depends on its speed ($\sigma_j$), reaction time ($t^*$), and deceleration ($\delta_j$). The length of each vehicle’s FSR is $c$ cells ($c$ is also called the longitudinal FSR cutoff value). Constant and identical deceleration is assumed for all non-ERVs (this can be relaxed in the future). The ‘ceil’ function in (1) returns the next higher integer value.

\[
MSD_j = \text{ceil} \left( \frac{t^* \sigma_j + 0.5 \frac{\sigma_j^2}{\delta_j}}{L} \right) \quad \text{(1)}
\]

\[
MFP_j = x'_j + MSD_j \quad \text{(2)}
\]

The optimization takes place on an assignment range (AR), downstream of the IR, that includes the FSR of each non-ERV. The AR has the same lateral size ($Y$) as the IR; however, its longitudinal size ($LL$) may be different. The AR’s starting longitudinal location ($AR_{start}$) depends on where the first FSR (among all the non-ERVs) is located along the link. As Figure 3-2 illustrates, the AR starts one increment before the smallest ($MFP_j$). Similarly, the AR’s ending longitudinal location depends on where the last FSR is located along the link. The AR ends at or after the highest possible longitudinal final position (i.e., maximum ($MFP_j + c$)). Hence, the AR’s longitudinal size ($LL$) is a multiple of the ERV size plus buffer, to obtain an integer number of ERV instructions in the AR. For each vehicle $j$, the minimum final longitudinal index ($x''_j$) with respect to the start of the AR is computed using (3).

\[
x''_j = MFP_j - AR_{start} + 1 \quad \text{(3)}
\]
3.3.4) Assumptions

In the proposed model, it is assumed that the speed, position and other physical characteristics of the ERV and the non-ERVs are available as inputs (e.g., from the connected vehicle environment or machine vision). It is assumed that no additional vehicles enter or leave the section during the ERV’s movement. We also assume that the ERV speed increases if a straight path is maintained while it decreases when the ERV performs a lane change, and that the speed depends on the non-ERV presence on cells adjacent to the ERV path (as discussed further near (26-31)). This paper is limited to a single ERV; for simplicity, ERVs are not indexed. Given the space limitation, this paper focuses on facilitating the ERV’s passage along a link segment and intersections are not considered.

3.4. Mathematical formulation

This formulation improves on initial work with nonlinear components [28]. Decision variables and parameters are described in Table 3-2 and Table 3-1.

Table 3-2: Variable notation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(w_{x,y})</td>
<td>Binary</td>
<td>ERV assignment variable that takes the value 1 if ERV is assigned to cell ((x, y)), 0 otherwise (i.e. 1 if the cell is part of ERV’s path during the time step)</td>
</tr>
<tr>
<td>(s^i)</td>
<td>Integer</td>
<td>ERV speed variable denoting the speed of the ERV at every increment (i)</td>
</tr>
<tr>
<td>(d_{k}^{x,y})</td>
<td>Binary</td>
<td>ERV instruction variable that takes the value 1 if the ERV is given instruction (k) at increment (i &lt; LL/(N+1)), and lateral position (y) ((k=1) means move right, (k=2) means go straight, (k=3) means move left)</td>
</tr>
<tr>
<td>(v^{x,y}_{j})</td>
<td>Binary</td>
<td>Non-ERV assignment variable that takes the value 1 if non-ERV (j) is assigned to cell ((x, y)) and 0 otherwise</td>
</tr>
<tr>
<td>(s_{Env}^{i})</td>
<td>Integer</td>
<td>ERV speed environment variable denoting the speed of the ERV only based on the ERV surrounding, computed at every increment (i &gt; 1)</td>
</tr>
<tr>
<td>(s_{temp}^{i})</td>
<td>Integer</td>
<td>Temporary ERV speed variable denoting the speed of the ERV based on the ERV surrounding and previous instruction but without accounting for minimum and maximum speed, computed at every increment (i &gt; 1)</td>
</tr>
<tr>
<td>(v^{i})</td>
<td>Binary</td>
<td>Variable that takes the value of 1 when (s_{temp}^{i} \geq S_{\text{min}}) and 0 otherwise, where (i &gt; 1)</td>
</tr>
</tbody>
</table>

Table 3-1: Parameter notation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Default Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(L, W)</td>
<td>n/a</td>
<td>Length/width of a cell in longitudinal/lateral direction</td>
</tr>
<tr>
<td>(N)</td>
<td>n/a</td>
<td>Number of longitudinal cells required to accommodate the ERV</td>
</tr>
<tr>
<td>(LL, Y)</td>
<td>n/a</td>
<td>Number of longitudinal/lateral cells in the AR</td>
</tr>
<tr>
<td>(AR_{\text{start}})</td>
<td>n/a</td>
<td>Distance (in cells) from the start of the Initial Range to the start of the Assignment Range</td>
</tr>
<tr>
<td>(\tau^j, \sigma_j, \delta_j)</td>
<td>2.5 s, n/a, 5 fps(^2)</td>
<td>Reaction time, initial speed in fps, and deceleration of vehicle (j)</td>
</tr>
<tr>
<td>(x^j_i, y^j_i)</td>
<td>n/a</td>
<td>Longitudinal index and lateral index of vehicle (j) in the IR</td>
</tr>
<tr>
<td>(b_{ij})</td>
<td>Binary</td>
<td>Binary parameter that takes the value of 1 if (y^j \geq y^i), and 0 otherwise</td>
</tr>
<tr>
<td>(MSD_{\text{mFP}}_{ij})</td>
<td>n/a</td>
<td>Minimum stopping distance of vehicle (j) and minimum final position of vehicle (j) (in cells)</td>
</tr>
<tr>
<td>(x^j_{\text{Min}})</td>
<td>n/a</td>
<td>Minimum final longitudinal index of vehicle (j) in cells with respect to the start of the AR</td>
</tr>
<tr>
<td>(c)</td>
<td>2</td>
<td>Longitudinal FSR cutoff value (in cells)</td>
</tr>
<tr>
<td>(S_{\text{min}}, S_{\text{free}})</td>
<td>1, n/a</td>
<td>Minimum and maximum ERV speed (in speed stage)</td>
</tr>
<tr>
<td>(M)</td>
<td>99999</td>
<td>Large number used to apply the Big M method</td>
</tr>
</tbody>
</table>
3.4.1) Objective Function

The objective function (4) maximizes the ERV’s speed while traveling within a link and the number of free cells adjacent to its path. The \((s^i)\) component is constrained by the previous instruction, the presence of surrounding non-ERVs during the last movement and the minimum and maximum ERV speeds. The \((s_{env}^i)\) component is only constrained by the presence of surrounding non-ERVs. Maximizing the summation of \((s_{env}^i)\) causes non-ERVs to be positioned further away from the ERV’s path to increase safety from human errors. [If the summation of \((s^i)\) were the only component of (4), the non-ERVs would be positioned as far as possible from the ERV path when the ERV maintains a straight path with a speed below the maximum \((S_{free})\). When the ERV changes lanes or when the speed cannot increase due to \(S_{free}\), the non-ERVs are not guaranteed to provide the ERV path with unoccupied adjacent cells, even if it were physically possible.]

The ERV speed variables are discrete but cannot be considered in cells per unit time since a discrete speed increase or decrease within an increment (short distance) would result in unrealistic acceleration/deceleration. So, the ERV speed variables \((s^i)\), \((s_{env}^i)\) and \((s_{temp}^i)\) are considered as integers and expressed in speed stages. Conversion from speed stage to speed (mph) depends on the ERV size \((N)\), ERV acceleration/deceleration capabilities, and cell length \((L)\) and is further discussed in the Appendix. Note that \(LL\) represents the longitudinal size of the AR and that \((\alpha_1)\) and \((\alpha_2)\) are the weights attributed to each term (see Section 3.6.1) 9).

\[
Max z = \alpha_1 \sum_{i=2}^{LL/(N+1)} s^i + \alpha_2 \sum_{i=2}^{LL/(N+1)} s_{env}^i \quad (4)
\]

3.4.2) Constraints

The constraints ensure that the ERV motion, ERV instructions, and non-ERV’s final positions are coordinated.

1) One vehicle per cell

Only one vehicle can occupy any given cell. At each cell \((x, y)\), the sum of the ERV assignment variable \((w_{x,y}^i)\) and all the non-ERV assignment variables \((v_{j,x,y}^i)\) should be 0 or 1.

\[
w_{x,y}^i + \sum_{j=1}^{L} v_{j,x,y}^i \leq 1; \ \forall (x,y) \quad (5)
\]

2) Non-ERVs stop in the feasible stopping range

To reduce the conflict between vehicles and to improve the computation time, the feasible space to which each of the non-ERVs can be assigned is reduced by setting a longitudinal FSR cutoff value \((c)\) beyond the minimum requirement. As indicated in Equation (6), each non-ERV is assigned to one cell in
its corresponding FSR (i.e., cells with a longitudinal index between \(x_j''\) and \(x_j'' + c\)), while Equation (7) ensures that each non-ERV is assigned to exactly one cell in the AR.

\[
\sum_{x=x_j''}^{x_j''+c} \sum_{y=1}^{y} v_j^{x,y} = 1; \quad \forall j \\
\sum_{x=1}^{\frac{LL}{2}} \sum_{y=1}^{y} v_j^{x,y} = 1; \quad \forall j
\]  

(6) (7)

3) **No passing among non-ERVs**

To reduce conflicts between vehicles, passing among non-ERVs is not allowed, as indicated in (8). If vehicle \(j\) is assigned to cell \((x, y)\), vehicle \(j'\) which was initially located downstream of \(j\) in the IR cannot stop at a cell upstream of \(j\) in the AR (i.e., \(j'\) cannot stop at a cell with a longitudinal index \(< x\)). Note that \(M\) is a large number (equal to 99999 here) used to apply the Big M method that allows the linearization of the constraint.

\[
\sum_{x=2}^{\frac{LL}{2}} \sum_{y=1}^{y} v_{j'}^{x,y} \leq M(1 - \sum_{y=1}^{y} v_j^{x,y}); \quad \forall x \geq 2; \forall j' > j
\]

(8)

4) **No weaving among non-ERVs**

Lateral conflicts (weaving) between vehicles are limited using (9-11). If vehicles \(j\) and \(j'\) are positioned on the same \(x\) index and \(j'\) is greater than \(j\), according to the labeling, vehicle \(j'\) is on the left of vehicle \(j\) in the IR. If vehicle \(j\) is assigned to cell \((x, y)\), vehicle \(j'\) can only stop on the left of or in front of \(j\) (i.e., at a cell with a lateral index \(\geq y\)), as indicated in (9).

\[
\sum_{x=1}^{\frac{LL}{2}} \sum_{y=1}^{y} v_{j}^{x,y} \leq M\left(\left|\frac{y_j'}{y'_j}\right| - 1\right) + M\left(1 - \sum_{x=1}^{\frac{LL}{2}} v_j^{x,y}\right); \quad \forall y; \forall j' > j
\]

(9)

If vehicles \(j\) and \(j'\) are not initially on the same \(x\) (i.e., \(j'\) is downstream of \(j\)) nor on the same \(y\) and if the final assignment places them on the same \(x\), they should be laterally positioned in the same way as in the IR, as indicated in (10) and (11). Note that \((y_j')\) is the initial lateral index of vehicle \(j\) in the IR. The binary parameter \((b_{jj'})\) takes the value of 1 when \((y_j' \geq y_j')\) and 0 otherwise. For example, if vehicle \(j'\) is on the left of \(j\) in the IR (i.e., \(y_j' > y_j'\)) then \((b_{jj'} = 0)\), and if vehicle \(j\) is assigned to cell \((x, y)\) in the AR, vehicle \(j'\) can stop at a cell with the same longitudinal index \(x\) only if the lateral index is strictly greater than \(y\). According to constraints (10), \(j'\) cannot stop on cells with longitudinal index \(x\) and lateral index less than \(y\). (In this case (11) is not a binding constraint.)

If vehicles \(j\) and \(j'\) have different longitudinal positions in the AR but the same lateral index \(y\) (i.e. \(y_j' = y_j'\)), no lateral constraint is needed. If vehicle \(j'\) stops on the same final longitudinal index as \(j\), it can
either stop on the right or left of vehicle $j$ (no preference). Otherwise, vehicle $j'$ stops at a position downstream of $j$ (due to the no passing constraint); in this case, the vehicles trajectories do not conflict.

\[
(1 - b_{jj'}) \times \left( \sum_{y'=1}^{y-1} v_{j'}^x v_{j'}^y \right) \times \left( |y' - y| \right) \leq M(1 - v_j^x v_j^y);
\]

\forall x; \forall y; \forall j' > j \quad (10)

\[
(b_{jj'}) \times \left( \sum_{y'=y+1}^{y} v_{j'}^x v_{j'}^y \right) \times \left( |y' - y| \right) \leq M(1 - v_j^x v_j^y);
\]

\forall x; \forall y; \forall j' > j \quad (11)

5) ERV passing lane

A minimum of one empty cell at every $x$ is reserved for the ERV, as shown in (12).

\[
\sum_{y=1}^{y} \sum_{j=1}^{j} v_j^x v_j^y \leq Y - 1; \forall x
\]

(12)

6) ERV instruction constraints

Only one set of instructions is sent to the ERV at each increment $i < LL/(N + 1)$, as indicated in (13).

\[
\sum_{y=1}^{y} \sum_{k=1}^{3} d_k^i \leq 1; \forall i = 1, ..., LL/(N + 1) - 1
\]

(13)

The ERV cannot move in a direction where no cell is available. This is reflected in (14) for ‘Move Right’ or $k=1$ and in (15) for the ‘Move left’ or $k=3$ instruction.

\[
d_1^i = 0; \forall i
\]

(14)

\[
d_3^i = 0; \forall i
\]

(15)

7) Setting the ERV instruction variables

The ERV’s continuous longitudinal motion is ensured by (16). Constraints (17-22) link the ERV assignment and ERV instruction variables and ensure continuous lateral motion. For example, if the ERV occupies a cell $(x = i(N + 1), y)$ and is instructed to move to the right lane at this cell then $(w_i^{(N+1), y} = 1)$ and $(d_1^i = 1)$. Using (17) and (18), the ERV assignment variable of each of the $(N + 1)$ downstream cells with lateral index $(y - 1)$ will be set to 1. Similarly, when the ‘Go straight’ or ‘Move left’ instructions are applied, the ERV assignment variables of the corresponding cells are set to 1.

\[
\sum_{y=1}^{y} w_j^x v_j^y = 1; \forall x
\]

(16)

\[
d_1^i \leq \frac{w_i^{(N+1), y} + w_i^{(N+1)+t,y-1}}{2}; \forall i = 1, ..., LL/(N + 1) - 1; \forall t = 1, 2, ..., N + 1; \forall y > 1
\]

(17)

\[
d_1^i \geq w_i^{(N+1), y} + w_i^{(N+1)+t,y-1} - 1; \forall i = 1, ..., LL/(N + 1) - 1; \forall t = 1, 2, ..., N + 1; \forall y > 1
\]

(18)
\[d^i_2 \leq \frac{w_i^{(N+1)\,y} + w_i^{(N+1)\,t,\,y}}{2}; \quad \forall i = 1, \ldots, LL/(N + 1) - 1; \forall t = 1, 2, \ldots, N + 1; \forall y \quad (19)\]
\[d^i_2 \geq w_i^{(N+1)\,y} + w_i^{(N+1)\,t,\,y} - 1; \quad \forall i = 1, \ldots, LL/(N + 1) - 1; \forall t = 1, 2, \ldots, N + 1; \forall y \quad (20)\]
\[d^i_3 \leq \frac{w_i^{(N+1)\,y} + w_i^{(N+1)\,t,\,y} + 1}{2}; \quad \forall i = 1, \ldots, LL/(N + 1) - 1; \forall t = 1, 2, \ldots, N + 1; \forall y < Y \quad (21)\]
\[d^i_3 \geq w_i^{(N+1)\,y} + w_i^{(N+1)\,t,\,y} + 1 - 1; \quad \forall i = 1, \ldots, LL/(N + 1) - 1; \forall t = 1, 2, \ldots, N + 1; \forall y < Y \quad (22)\]

8) Setting the non-ERV assignment variables

Constraints (23-25) ensure that cells which are part of the ERV’s path are empty and not occupied by non-ERVs. For forward motion (23), it is assumed that the ERV needs \((N + 1)\) cells in the same \(y\) open. Thus, when \((d^i_2)\) is equal to 1 (i.e., ERV instructed to ‘Go straight’), the sum of all the non-ERV assignment variables of the cells with a longitudinal index between \((x + 1)\) and \((x + (N + 1))\) and lateral index \(y\) should be equal to 0. For lane changing (24) and (25), the ERV is assumed to need \((N)\) forward cells in the same \(y\) to be able to maneuver and move to the adjacent lane in which \((N + 1)\) forward cells are free.

\[\sum_{x=t'}^{T'} \sum_{j=1}^{J} v^x_{j} \leq M(1 - d^i_2); \quad \forall y; \forall i = 1, \ldots, LL/(N + 1) - 1;\]
\[t' = (N + 1)i + 1; T' = (N + 1)i + (N + 1) \quad (23)\]
\[\sum_{x=t''}^{T''} \sum_{j=1}^{J} v^x_{j} + \sum_{x=t'}^{T'} \sum_{j=1}^{J} v^x_{j} \leq M(1 - d^i_1); \quad \forall y > 1;\]
\[\forall i = 1, \ldots, LL/(N + 1) - 1; t' = (N + 1)i + 1; T' = (N + 1)i + (N + 1); T'' = (N + 1)i + N \quad (24)\]
\[\sum_{x=t'}^{T'} \sum_{j=1}^{J} v^x_{j} + \sum_{x=t''}^{T''} \sum_{j=1}^{J} v^x_{j} \leq M(1 - d^i_3); \quad \forall y < Y;\]
\[\forall i = 1, \ldots, LL/(N + 1) - 1; t' = (N + 1)i + 1; T' = (N + 1)i + (N + 1); T'' = (N + 1)i + N \quad (25)\]

9) ERV Speed Constraints

In this formulation, three speed variables are used \((s^i, s_{env}^i, s_{temp}^i)\). To determine the ERV speed \((s^{i+1})\) at increment \((i + 1)\), \((s_{env}^{i+1})\) and \((s_{temp}^{i+1})\) are first identified based on the ERV speed \((s^i)\) at the previous increment \((i)\).

While traveling, the ERV adjusts its speed based on its surroundings. A free adjacent lane (no non-ERVs) enables the ERV to increase its speed while occupied adjacent lanes make its movement slower. As shown in (26-31), a speed \((s_{env}^i)\) is assigned at each increment and is only constrained by the number of nearby stopped vehicles. The summation of \((s_{env}^i)\) is maximized in (4) to ensure that the non-ERVs are
positioned as far as possible from the ERV path when the ERV is performing a lane change or when the ERV speed is at the maximum ($S_{free}^\text{env}$).

If the ERV is in a middle lane (26-27), and its right and left downstream cells are unoccupied, the ERV can increase its speed by 1. If one side is occupied, the speed remains constant. If both sides are occupied, the ERV’s speed decreases by 1. If the ERV is positioned in the rightmost lane (28-29) and its left side is unoccupied, the ERV speed can increase by 1. Otherwise, the speed remains constant. Constraints (30-31) present the analogous situation for the ERV positioned in the leftmost lane ($y = Y$).

Separate constraints (27), (29) and (31) were introduced for the last increment ($i = LL/(N + 1)$) due to the end of the AR that limits the ERV surrounding space affecting the ERV speed environment after the last movement ($s_{env}^{LL/(N+1)}$).

$$s_{env}^{i+1} \leq s^i + 1 - \sum_{j=1}^{f} v_j^{(N+1)i+\tau_y-1} - \sum_{j=1}^{f} v_j^{(N+1)i+\tau_y+1} + M(1 - w^{(N+1)i+\tau_y});$$

$$\forall i = 1, \ldots, LL / N + 1 - 2; \forall t = 0,1,\ldots,N + 2; \forall y = 2,3,\ldots,Y - 1 \quad (26)$$

$$s_{env}^{LL/(N+1)} \leq s^{LL/(N+1)-1} + 1 - \sum_{j=1}^{f} v_j^{LL-(N+1)+\tau_y-1} - \sum_{j=1}^{f} v_j^{LL-(N+1)+\tau_y+1} + M(1 - w^{LL-(N+1)+\tau_y});$$

$$\forall t = 0,1,\ldots,N + 1; \forall y = 2,3,\ldots,Y - 1 \quad (27)$$

$$s_{env}^{i+1} \leq s^i + 1 - \sum_{j=1}^{f} v_j^{(N+1)i+\tau_2} + M(1 - w^{(N+1)i+\tau_2});$$

$$\forall i = 1, \ldots, LL/(N + 1) - 2; \forall t = 0,1,\ldots,N + 2 \quad (28)$$

$$s_{env}^{LL/(N+1)} \leq s^{LL/(N+1)-1} + 1 - \sum_{j=1}^{f} v_j^{LL-(N+1)+\tau_2} + M(1 - w^{LL-(N+1)+\tau_2});$$

$$\forall t = 0,1,\ldots,N + 1 \quad (29)$$

$$s_{env}^{i+1} \leq s^i + 1 - \sum_{j=1}^{f} v_j^{(N+1)i+\tau_y-1} + M(1 - w^{(N+1)i+\tau_y});$$

$$\forall i = 1, \ldots, LL/(N + 1) - 2; \forall t = 0,1,\ldots,N + 2 \quad (30)$$

$$s_{env}^{LL/(N+1)} \leq s^{LL/(N+1)-1} + 1 - \sum_{j=1}^{f} v_j^{LL-(N+1)+\tau_y-1} + M(1 - w^{LL-(N+1)+\tau_y});$$

$$\forall t = 0,1,\ldots,N + 1 \quad (31)$$

A temporary variable ($s_{temp}^i$) is assigned at each increment and is constrained by the surrounding conditions (32) and the instruction given at the previous increment (33). The assumption is that the ERV can increase its speed by 1 if it is going straight. However, if it moves right or left, its speed decreases by 1.
\[ s_{\text{temp}}^{i+1} \leq s_{\text{env}}^{i+1}; \ \forall \ i = 1, ..., LL/(N + 1) - 1 \]  

\[ s_{\text{temp}}^{i+1} \leq s^i + 2 \sum_{y=1}^{N} d_y^{i+1} - 1; \ \forall \ i = 1, ..., LL/(N + 1) - 1 \]  

The ERV speed \( s^{i+1} \) is the actual speed that can be adopted by the ERV and is limited by the maximum allowable ERV speed \( S^{\text{free}} \), as shown in (34).

\[ s^{i+1} \leq S^{\text{free}}; \ \forall \ i = 1, ..., LL/(N + 1) - 1 \]  

Since the speed of the ERV \( s^{i+1} \) cannot decrease below the minimum speed \( S^{\text{min}} \), \( s^{i+1} \) should be equal to the minimum of \( s_{\text{env}}^{i+1} \) and \( s^i + \text{(instruction factor)} \) if their minimum is greater than \( S^{\text{min}} \). Otherwise, \( s^{i+1} \) should take the value of \( S^{\text{min}} \). In other words, \( s^{i+1} \) should be equal to the maximum of \( (S^{\text{min}}) \) and \( (s_{\text{temp}}^{i+1}) \), as reflected in (35-38). Note that \( (s_{\text{temp}}^{i+1}) \) is primarily introduced for practical purposes, as a temporary ERV speed variable taking the minimum of \( s_{\text{env}}^{i+1} \) and \( s^i + \text{(instruction factor)} \) (i.e., right hand side of (33)).

\[ s^{i+1} \geq S^{\text{min}}; \ \forall \ i = 1, ..., LL/(N + 1) - 1 \]  

\[ s^{i+1} \geq s_{\text{temp}}^{i+1}; \ \forall \ i = 1, ..., LL/(N + 1) - 1 \]  

\[ s^{i+1} \leq M(1 - v^{i+1}) + s_{\text{temp}}^{i+1}; \ \forall \ i = 1, ..., LL/(N + 1) - 1 \]  

\[ s^{i+1} \leq Mv^{i+1} + S^{\text{min}}; \ \forall \ i = 1, ..., LL/(N + 1) - 1 \]  

10) Binary, integer and initial conditions constraints

Based on the variables’ type, integer and binary constraints are added along with constraints indicating the ERV’s initial lateral position and speed at the beginning of the AR.

3.5. Description of the experiment

The experimental analysis has two parts: sensitivity to initial parameters and comparison to current practice. The cell size is the regular vehicle size plus buffers (\( L = 21 \) feet and \( W = 10 \) feet).

3.5.1) Sensitivity analysis

For each test, a given parameter is varied while fixing the other parameters to default values shown in Table 3-3. The six base scenarios are described by road and ERV type: Arterial/Ambulance, Major

<p>| Table 3-3: Base case scenarios parameters and resulting ERV path description |
|------------------------|-----------------|----------------|-----------------|----------------|-------------------|-----------------|</p>
<table>
<thead>
<tr>
<th>ERV size</th>
<th>Road type</th>
<th>ERV initial speed (stage)</th>
<th>ERV initial lateral index</th>
<th>Road composition</th>
<th>Number of non-ERVs</th>
<th>Non-ERV initial position</th>
<th>Non-ERV speed (mph)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambulance (N=2 cells)</td>
<td>Arterial</td>
<td>8</td>
<td>3</td>
<td>1 shoulder and 4 lanes</td>
<td>15</td>
<td>Dispersed</td>
<td>Homogenous =40</td>
</tr>
<tr>
<td></td>
<td>Major collector</td>
<td>4</td>
<td>3</td>
<td>1 shoulder and 3 lanes</td>
<td></td>
<td></td>
<td>Homogenous =30</td>
</tr>
<tr>
<td></td>
<td>Minor collector</td>
<td>3</td>
<td>2</td>
<td>1 shoulder and 2 lanes</td>
<td></td>
<td></td>
<td>Homogenous =20</td>
</tr>
<tr>
<td>Police car (N=1 cells)</td>
<td>Arterial</td>
<td>6</td>
<td>3</td>
<td>1 shoulder and 4 lanes</td>
<td></td>
<td></td>
<td>Homogenous =40</td>
</tr>
<tr>
<td></td>
<td>Major collector</td>
<td>3</td>
<td>3</td>
<td>1 shoulder and 3 lanes</td>
<td></td>
<td></td>
<td>Homogenous =30</td>
</tr>
<tr>
<td></td>
<td>Minor collector</td>
<td>2</td>
<td>2</td>
<td>1 shoulder and 2 lanes</td>
<td></td>
<td></td>
<td>Homogenous =20</td>
</tr>
</tbody>
</table>
Collector/Ambulance, Minor Collector/Ambulance, Arterial/Police, Major Collector/Police, Minor Collector/Police. Speed limits are 55, 35, and 25 mph, and the number of lateral cells $Y$ is 5, 4, and 3 for arterials, major collectors, and minor collectors, respectively. The ERV minimum speed is 5 mph while its maximum speed is 10 mph above the speed limit. The speeds in mph are converted to the corresponding ERV stage based on the ERV type. (A sample speed-stage table is in the Appendix).

In the base scenarios, the ERV’s initial speed stage is in the middle of its range; the ERV is positioned on the middle lane and roads have only a right shoulder. The IR is ten cells (210 feet) long. For each base scenario, equal weights ($\alpha_1 = \alpha_2 = 1$) are assigned in the objective function.

3.5.2) Comparison to local practice

The formulation’s output is compared to the one obtained from the local practice “Go to the nearest edge” (rightmost or leftmost lane). Seven tests are executed: six for the base scenarios and one test with an ambulance positioned initially on the rightmost lane of a Major Collector.

3.6) Experimental analysis

The tests were executed using the CPLEX solver on the NEOS server (with CPU at 2.2-2.8 GHz and 64-192 GB RAM) [29-31]. CPLEX uses the branch-and-bound technique to optimize integer programs. In this formulation, the number of variables and constraints varies with the number and speeds of non-ERVs, the ERV size, and the AR’s length and geometry.

3.6.1) Sensitivity Analysis

1) ERV initial position

The ERV initial lateral position in the AR varies between $y=1$ and $y=Y$. As shown in Figure 3-3, when traveling on the widest tested road type (arterial), the ERV’s initial position does not affect the ERV speed. A straight movement, allowing a linear speed increase, results since the side(s) of the ERV can be free at all times. For narrower links (major or minor collectors), if positioned on an edge, the ERV maintains a straight path and the speed increases until reaching a plateau at $S_{\text{free}}$. If positioned on a middle lane, the ERV either continues straight or moves to one of the edge lanes. Even though a lane change decreases the ERV speed, it can improve the objective function value, as travelling on an edge necessitates freeing fewer cells to allow a speed increase than traveling on a middle lane. The narrower the link, the earlier the ERV changes to an edge lane.

2) ERV initial speed

The minimum and maximum ERV speed stages change with the ERV size and road type. This test varies the ERV initial speed between the corresponding minimum and maximum allowable ERV speeds. On arterials, for all initial speeds, the ERV maintains a straight path and increases its speed until reaching...
the maximum speed ($S_{free}$). This means both sides of the ERV can be freed at all downstream cells. When the ERV speed is initially $S_{free}$, it remains constant. On major collectors, when the initial speed is less than $S_{free}$, the ambulance maintains a straight movement and the speed increases until reaching a plateau due to the presence of non-ERVs on one of its sides. For the scenario with maximum initial speed, the ambulance “moving to the closest edge” and “maintaining a straight path and constant speed” are alternate optimal solutions. At all initial speeds, police cars move to the closest edge when “maintaining a straight path” will not result in a speed increase due to the presence of non-ERVs on the side of the ERV path. Moving to the nearest edge is preferred to “maintaining a straight path” due to the $s_{env}$ term in the objective function (see also Section 3.6.10). On minor collectors, and in all scenarios, both ERV types move to the nearest edge at the beginning of the AR. Once there, the speed increases until non-ERVs on cells adjacent to the ERV path force the speed to remain constant.

3) Road composition

Different road compositions dictate different non-ERV IR lateral positions since they are not initially on shoulders. A link with 4 lateral cells has 3 lanes and 1 shoulder or 2 lanes and 2 shoulders. With 3 lanes and 1 shoulder, non-ERVs are more dispersed. For different compositions with the same number of lateral cells, the same objective value, ERV path, ERV speed and non-ERV assignment (or alternate optimum) are obtained. In the tested scenarios, the output is insensitive to road composition for a given number of non-ERVs.

Figure 3-3: Variation of ERV speed in the AR with different ERV initial positions (per road type and ERV size). LC refers to “Lane Change”, so charts without “LC” indications represent a straight ERV path.
4) ERV size

Size affects the number of instructions and maneuvers that can be made in a given AR. Smaller ERVs can make more maneuvers, allowing them to achieve more speed increases.

5) Number of non-ERVs

To evaluate congestion effects, the base scenarios are tested with 10, 15 and 20 non-ERVs. When the ERV is traveling on an arterial, the output is insensitive to the tested increase in non-ERVs. The ERV follows the same straight path with free adjacent cells on both sides along the AR. With these demands, on arterials, the link is wide enough to free both sides of the ERV even with 20 non-ERVs in the AR. For narrower links, as the number of non-ERVs increases, the ERV speeds are reduced due to the positioning of non-ERVs on adjacent cells, and the ERV path involves lane changes.

6) Non-ERV initial position

For each ERV size and road type, three non-ERV initial positions are tested: dispersed, clustered at the beginning, and clustered at the end of the IR. Dispersed non-ERVs resulted in higher ERV speeds along the AR than the scenarios with initially clustered positions, for all road types and ERV sizes. With dispersed non-ERVs and homogenous speeds, the FSRs of the non-ERVs are dispersed in the AR allowing more effective use of the AR space and free ERV path sides.

With dispersed non-ERVs, the model positions the non-ERVs as far downstream as possible to allow the ERV to increase its speed before being forced to remain constant. When non-ERVs are clustered at the end of the IR, the non-ERVs’ FSRs are located toward the end of the AR. This is why (1) the scenario with non-ERV positions clustered at the end of the IR and the scenario with dispersed non-ERVs led to close results and (2) clustering non-ERVs at the end of the IR resulted in better ERV speeds than when they are clustered at the beginning, for all road types and ERV sizes. To improve the output for non-ERVs clustered at the beginning of the IR, the longitudinal FSR cutoff value \( c \) could be increased.

7) Heterogeneous non-ERV speed

In all previous tests, the non-ERV speeds are homogenous, and the longitudinal FSR cutoff default value \( c \) is 2 cells beyond the minimum final longitudinal index \( x_{j''} \). When random speeds are assigned to non-ERVs and \( c \) is 2, the formulation did not result in a feasible solution for all tested scenarios. For instance, suppose vehicle \( j' \) is downstream of vehicle \( j \). Vehicle \( j \), travelling with a higher speed, has a more downstream FSR than vehicle \( j' \) travelling at a lower speed. Since the formulation ensures that each vehicle is assigned to a cell within its corresponding FSR and passing is not allowed, no feasible solution is generated because the FSR of \( j' \) ends before the start of the FSR of \( j \). A larger \( c \) value extends the FSRs and
allows vehicle \( j' \) to travel to a cell at the same or higher longitudinal index than the final position of vehicle \( j \), resulting in a feasible solution.

On arterials (for both ERV sizes), \( c \) should increase from 2 to 22, while on major collectors it should be extended to 11 to obtain a feasible solution. On minor collectors, no adjustment is needed. As the link type gets narrower, the variance of the non-ERVs speeds gets smaller resulting in closer FSRs and less need for cutoff value extensions.

With homogenous non-ERV speeds, higher ERV speeds can be achieved since the AR space can be used effectively. With random speeds, some cells in the AR cannot be occupied by any non-ERVs since the cells are not in their FSRs.

8) Computation time

The average computation times increase as the road type becomes narrower, going from arterial (0.22 s) to major collector (0.24 s) to minor collector (0.32 s). As the road gets narrower (with fewer lateral cells and variables), the search for the optimal ERV path becomes more challenging given that the number of non-ERVs remains the same. If the ERV is initially positioned on middle lanes, it likely moves to an edge on narrower roads. Paths that include lane changes result in higher average computation times (0.34 s) than straight paths (0.21 s). When changing lanes, more downstream cells have to be freed to perform the maneuver, activating more constraints.

As the number of non-ERVs increases (10 - 20), the average computation time increases (0.33 - 1.33 s) due to the increased number of variables. In addition, with more congestion, right/left ERV maneuvers become more likely. Thus, the computation time is sensitive to the number of non-ERVs.

To examine the implications that further increases in the IR size and number of non-ERVs have on the computation times, the tests shown in Table 3- 4 were executed. The computation time obtained in Test 2 is relatively high (more than 2 minutes) given that the non-ERV speeds are homogenous and the non-ERVs are dispersed in the IR (see above). In Test 1, if the non-ERVs are traveling at a homogenous speed of 40 mph at the time of data collection, after the computation time of 25.42 s, they would have traveled

<table>
<thead>
<tr>
<th>Table 3- 4: Limiting scenarios input/output data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input</strong></td>
</tr>
<tr>
<td>Road type</td>
</tr>
<tr>
<td>ERV Size</td>
</tr>
<tr>
<td>Number of non-ERVs</td>
</tr>
<tr>
<td>IR Longitudinal size</td>
</tr>
<tr>
<td><strong>Output</strong></td>
</tr>
<tr>
<td>AR Longitudinal size</td>
</tr>
<tr>
<td>Computation time</td>
</tr>
</tbody>
</table>
approximately 1500 ft. The initially retrieved non-ERV positions should be adjusted to reflect the location at which the non-EVRs receive the assignments.

9) Objective Function Weight Analysis

The objective function is the summation of two elements with the same units. Initially, equal weights \((\alpha_1 = \alpha_2 = 1)\) were assigned. To evaluate the impact that weight combinations have on the output, the six base scenarios were tested with the weights: \((\alpha_1 = 2, \alpha_2 = 1); (\alpha_1 = 1, \alpha_2 = 1); (\alpha_1 = 1, \alpha_2 = 2)\).

As shown in Table 3-5, on arterials, the same ERV straight paths were observed since there is enough space to achieve the highest values of \(s^i\) and \(s_{env}^i\). On minor collectors, due to limited space, moving to the edge is better; even if \(s^i\) decreases at the increment after the lane change, \(s^i\) and \(s_{env}^i\) will then be able to reach higher values by freeing one side of the ERV while, if the ERV remains on the middle lane and one side is free, the \(s_{env}^i\) and subsequently \(s^i\) values will not increase. On major collectors, with the different weight sets, different outputs were obtained. As more weight is attributed to element (1) \((\alpha_1 > \alpha_2)\), a straight path results. However, as more weight is given to element (2) \((\alpha_2 > \alpha_1)\), a lane change (to the edge) that can achieve higher \(s_{env}^i\) occurs. Equal weights once generated a solution like weight set (1) and once like weight set (3). This unbiased weight set was used in the sensitivity analysis tests. Selecting weights is subjective. If keeping the ERV further away from non-ERVs is preferred, more weight should be assigned to element (2) in the objective function.

Table 3-5: Objective function weight analysis results

<table>
<thead>
<tr>
<th>Base case scenarios</th>
<th>Weights</th>
<th>Ambulance</th>
<th>Police car</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\alpha_1)</td>
<td>(\alpha_2)</td>
<td>Arterial</td>
</tr>
<tr>
<td>(1)</td>
<td>2</td>
<td>1</td>
<td>Straight ERV path</td>
</tr>
<tr>
<td>(2)</td>
<td>1</td>
<td>1</td>
<td>Straight ERV path</td>
</tr>
<tr>
<td>(3)</td>
<td>1</td>
<td>2</td>
<td>Lane change in ERV path</td>
</tr>
</tbody>
</table>

3. 6. 2) Comparison to local practices

Local practices attempt to reduce confusion by providing simple rules, such as stopping at the nearest edge when an ERV approaches [32]. Passing and weaving between vehicles may occur as each driver acts independently. To compare the ERV path and speeds that could be generated with the local practice “Go to the nearest edge” and the solution generated by the proposed formulation, the six base scenarios are tested with both approaches. On arterials, the ERV paths and speeds were identical. Even though no speed benefits are observed on arterials, the formulation could eliminate confusion, as well as passing and weaving among vehicles, improving their safety.
When traveling along a major or minor collector, the speed benefits of the proposed formulation are substantial. For local practice, due to the positioning of non-ERVs along the edges, and as the road becomes narrower, the ERV (initially positioned on a middle lane) travels on cells that have adjacent occupied cells, resulting in a speed plateau or decrease. However, the proposed model pushes the non-ERVs away from the ERV, to allow the ERV’s speed to increase. When non-ERVs are positioned according to the formulation, the ERV reaches greater speeds at the end of the AR (Table 3-6).

<table>
<thead>
<tr>
<th>Speeds (in stage)</th>
<th>Major Collector</th>
<th>Minor Collector</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ambulance</td>
<td>Police car</td>
</tr>
<tr>
<td><strong>Initial speed</strong></td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td><strong>Final speed</strong></td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

In a scenario in which an ambulance is initially positioned on the right edge of a major collector, under local practice, non-ERV drivers may move to the edge before identifying the lane on which the ERV is traveling. The ERV may be forced to change lanes to avoid the non-ERVs on the same lane. As shown in Figure 3-4, using the proposed formulation, the final non-ERV positions are optimized and the ERV remains on the same lane, increases its speed and reaches the maximum allowable speed. In this scenario, the ERV needs 4.54 seconds to travel the 252-feet link segment (average speed of 37.85 mph) in which the non-ERVs’ final positions are optimized while it needs 6.78 seconds to travel the same link segment when the non-ERVs stop at the nearest edge (average ERV speed of 25.34 mph). This travel time improvement on a relatively short segment (252 feet) is promising. As a link becomes narrower, shorter travel times are observed when using the proposed formulation. Specifically, the travel time improvements observed when an ambulance is initially positioned on the right edge of a minor collector are more significant; the average ERV speed with optimized non-ERV positions is 31.94 mph compared to ERV speed of 7.48 mph when non-ERV positions are not optimized. Additional figures are available from the authors upon request. Travel time improvements for larger transportation link segments will be discussed in future research as strategies will be adopted to control the timing overhead resulting from the increase in the problem size.
Summary and conclusions

This paper presented an ILP to assist an ERV by delineating its intra-link path and providing maneuvering instructions at every increment. It limits the confusion experienced by non-ERVs by assigning each of them a position along the link depending on the non-ERV’s feasible stopping distance. The ILP’s objective is to maximize the ERV’s speed and the free space surrounding its path. The program can adapt to various ERV characteristics, road types and other parameters.

A sensitivity analysis was conducted to evaluate the impact of varying the main parameters on the output and computation time. On narrower roads, as the number of non-ERVs increased and they were more clustered, the ERV’s speed was less likely to increase, and its path involved a lane change in cases when the ERV was initially positioned on a middle lane. Given the same AR length, scenarios with a narrower road, a larger number of clustered non-ERVs and whose output ERV path involved lane changes required longer computation times. Yet, the computation times were small in the examined scenarios with an average of 0.46 s (for an AR 15 cells long for ERV size =2 and 14 cells long for ERV size =1) and were the most sensitive to the number of non-ERVs. If non-ERVs have random speeds, larger longitudinal FSR cutoff values may be required to obtain a feasible solution. The proposed formulation generated higher ERV speeds than current practice requiring non-ERVs to move to the nearest edge, especially when the link is narrow and in scenarios where the ERV is initially on an edge. On wide links, current practice and this formulation led to the same results. The advantages of the ILP, in this case, are reduced confusion, passing/weaving among vehicles, and other safety issues.

In the future, this ILP will be extended to a complete link (multiple segments) and intersections. With the initial benefits on a small segment, more notable travel time savings are anticipated over the complete journey. In addition, some assumptions will be relaxed, and simulations will be used to for further evaluation and to test market penetration levels.

Figure 3-4: Variation of ERV speed in the AR with different practices (Major Collector/Ambulance)
Appendix A

In this mathematical formulation, the ERV speed variables are integers expressed in speed stages and not in units of distance per time. Based on our approach, the ERV speed can either increase or decrease within an increment (fixed, short distance). If the speed were expressed in units of distance per time with continuous values, maintaining a straight path with free adjacent cells when traveling at lower speeds would be prioritized compared to travelling at higher speeds. (When traveling at lower speeds, the ERV requires a longer time to travel a fixed distance. If it were accelerating with a uniform magnitude for a longer time, a greater speed change would result, and thus a greater improvement to the objective function, compared to what would result from higher speeds.) To assign the same priority to all ERV speeds, we assume that the ERV speed variables can increase or decrease by one unit within an increment and hence take integer values. Furthermore, if the ERV speeds variables were expressed in one of the common units of speeds (such as mph, ft/sec and km/h), to achieve an increase or decrease of X units of speed within an increment (short distance), unrealistic accelerations and decelerations would arise (very small or very large magnitudes depending on X, the unit of speed adopted, and the initial ERV speed). This is the second reason why the ERV speed variables are expressed in speed stages and may increase or decrease by one unit within an increment. Moving to a higher stage means increasing the speed but with a practical acceleration. To identify the ERV speed in distance per time at each speed stage, a preprocessing step that consists of developing a lookup table corresponding to each ERV type and road type is executed as follows: First, the following parameters are identified: (1) minimum and maximum allowable ERV speeds based on the roadway type and the ERV type, (2) cell length \( L \), (3) ERV longitudinal size \( N \) (in cells), and (4) ERV acceleration capabilities \( a \). Second, the speed (in distance per time) corresponding to each stage is computed. The speed at stage 1 is simply the minimum allowable ERV speed. The speed at stage \( g \) (\( g > 1 \)) is computed as follows:

- Find the time \( t \) needed to travel the distance \( d = (N + 1) \times L \) with an initial speed \( V_i \) (the speed at stage \( g-1 \)) and acceleration \( a \): \( d = V_i t + \frac{1}{2} at^2 \).
- Find \( V_f \) (the speed at stage \( g \)) after the elapsed time \( t \):
  \[
  V_f = V_i + at.
  \]

Third, stages are added until the maximum stage is reached.

The lookup table corresponding to each ERV type is developed using a cell size \( L=21 \) ft, ERV longitudinal size \( N=2 \) (ambulance) and \( N=1 \) (police car) and acceleration/ deceleration = 5 ft/s² (for ambulance) and 10 ft/s² (for police). The ERV minimum speed is 5 mph while its maximum speed is 10 mph above the speed limit. The lookup table corresponding to an ambulance is shown below:
<table>
<thead>
<tr>
<th>ERV type: Ambulance (N=2)</th>
<th>ERV speed stage</th>
<th>Road type</th>
<th>ERV speed mph</th>
<th>Min speed</th>
<th>Mid speed</th>
<th>Max speed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>Arterial</td>
<td>5.00</td>
<td>Min speed</td>
<td>Mid speed</td>
<td>Max speed</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Major Collector</td>
<td>17.83</td>
<td>Min speed</td>
<td>Mid speed</td>
<td>Max speed</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Minor Collector</td>
<td>24.71</td>
<td>Min speed</td>
<td>Mid speed</td>
<td>Max speed</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Min speed</td>
<td>30.06</td>
<td>Mid speed</td>
<td>Max speed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Min speed</td>
<td>34.59</td>
<td>Max speed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Min speed</td>
<td>38.59</td>
<td>Max speed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Mid speed</td>
<td>42.22</td>
<td>Max speed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>Mid speed</td>
<td>45.55</td>
<td>Max speed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>Mid speed</td>
<td>48.66</td>
<td>Max speed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>Mid speed</td>
<td>51.58</td>
<td>Max speed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>Mid speed</td>
<td>54.35</td>
<td>Max speed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>Mid speed</td>
<td>56.98</td>
<td>Max speed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>Mid speed</td>
<td>59.49</td>
<td>Max speed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>Max speed</td>
<td>61.91</td>
<td>Max speed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>Max speed</td>
<td>64.23</td>
<td>Max speed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>Max speed</td>
<td>66.47</td>
<td>Max speed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Mid ERV speed: initial ERV speed in base case scenarios*

**Reference**


Chapter 4. SEQUENTIAL OPTIMIZATION OF AN EMERGENCY RESPONSE VEHICLE’S INTRA-LINK MOVEMENT ALONG TRANSPORTATION SEGMENTS IN A PARTIALLY CONNECTED VEHICLE ENVIRONMENT

This paper introduces a semi-automated system that facilitates Emergency Response Vehicle (ERV) movement through a transportation link by providing instructions to downstream non-ERVs. The proposed system adapts to information from non-ERVs that are nearby and downstream of the ERV. As the ERV passes stopped non-ERVs, new non-ERVs are considered. The proposed system sequentially executes integer linear programs (ILPs) on transportation link segments with information transferred between optimizations to ensure ERV movement continuity. This paper extends a previously developed mathematical program that was limited to a single short segment. The new approach limits runtime overhead without sacrificing effectiveness and is more suitable to dynamic systems. It also accommodates partial market penetration of connected vehicles using a heuristic reservation approach, making the proposed system beneficial in the short-term future. The system accommodates unconnected vehicles by estimating their positions and then allowing connected non-ERVs to slightly adjust their instructed final positions while ensuring that the optimal ERV movement is unaffected. The proposed system can also assign the ERV to a specific lateral position at the end of the link, a useful capability when next entering an intersection. Experiments were conducted to develop recommendations on link segment sizes and grouping of non-ERVs in the different ILP optimizations to reduce computation times without compromising efficiency. Large gaps between non-ERVs should act as delimiters between link segments. When compared to the current practice of moving to the nearest edge, the system reduces ERV travel time an average of 3.26 seconds per 0.1 mile and decreases vehicle interactions.

4.1. Introduction

Emergency Response Vehicles (ERVs) face numerous challenges when navigating to and from emergencies. Downstream vehicles should coordinate their movements to cooperatively open a path for the ERV to reach its destination as quickly as possible and safely. ERV preemption is a technique activated at traffic signals upon the detection of an approaching ERV by direct microwave receipt, optical recognition, pavement loops, etc. and provides an earlier or extended green time to the ERV’s approach [1]. ERV preemption adjusts the signal timings so that approaching ERVs do not wait in a queue at red signals and do not collide with vehicles entering the intersections from the opposing approaches.

With advancements in wireless communication technologies and hardware/software in-the-loop techniques, ERV preemption has continuously evolved by integrating new methods to autonomously monitor the real-time status of a signalized intersection [2] and to identify the best transition phase that
minimizes the impact to other vehicles after each preemption [3-7]. Conversely, the challenges experienced when travelling along transportation links need more investigation as the current system is still limited to visual and audible warning systems (i.e., lights and sirens) that often fail to prevent confusion and vehicle collisions [8, 9], especially for police vehicles [9].

Hannoun et al. [10] presented a mathematical program to facilitate ERV passage along a transportation link of predefined length relying on vehicular communications, assuming all vehicles are connected. The integer linear program (ILP) presented generates the fastest ERV intra-link path along with the ERV’s maneuvering instructions and speeds based upon collected downstream information. The ILP also assigns each downstream non-ERV to a specific position along the link. In [10], the proposed system is limited to short links so it is only sufficient if activated by the ERV driver needing assistance over a short distance. However, the ERV’s driver may request extended assistance, especially in congested urban transportation networks. To practically facilitate the passage of the ERV along its complete path from origin to destination, an updated approach requires solutions to the following challenges left unaddressed from the previous methodology: (1) substantial computation time due to the considerably large problem size (2) unnecessary early notification and instruction dissemination to non-ERVs travelling downstream on the ERV path and distant from the ERV position.

In this paper, the problem size limitation is tackled and the facilitation of the ERV’s movement along larger links and along its complete trip from origin to destination is made possible. With downstream information becoming available and updated as the ERV is travelling and approaching its destination, a sequential optimization approach that consists of applying the ILP proposed in [10] on multiple link segments consecutively is proposed. This extends the approach so that the subsequent application of the modified ILP considers the output from the application to the previous segment, which may have already assigned non-ERVs to locations that are included in the space being currently optimized. Hence, the extensions presented in this paper act as crucial steps to achieving applicability of the ILP optimization on complete transportation networks.

This study relies on connected vehicle technologies: (1) for the collection of the positions and speeds of the vehicles required as input to the ILP and (2) for the dissemination of the instructions to ERVs and non-ERVs. Although the connected vehicle environment is rapidly evolving, the deployment of its hardware and software components is challenging. The market penetration rate of connected vehicles will remain low in the near-term future [11], however it is expected that a near 100% penetration rate will only be observed in about 25 to 30 years [12], which prevents the work in [10] from being directly applicable in the near future. In [10], all non-ERVs are expected to share their data and receive their corresponding instructions which is a major limitation for short-term deployment. In this paper, the assumption of full market penetration is resolved by introducing a technique that estimates the presence of unconnected
vehicles and reserves additional spaces for the unconnected vehicles. It is important to assess the functionality and performance of connected vehicle applications under different market penetrations [13].

The remainder of this paper is divided into six sections. Section 4. 2. briefly reviews previous studies about the enhancement of emergency response vehicles’ operations and techniques to mitigate the partial market penetration in connected vehicle applications. Section 4. 3. presents the system proposed in this paper along with its assumptions. Section 4. 4. outlines the preprocessing steps required prior to each optimization. The adjusted ILP formulation (objective function along with old and new constraints) is then introduced in Section 4. 5. Section 4. 6. includes the experimental plan and results from the evaluation of the system with partial market penetration and from comparison with current practices. Finally, Section 4. 7. presents key conclusions and suggestions for future work.

4. 2. Literature Review

The connected vehicle environment allows the development of new applications focusing on enhancing safety, mobility, and/or the environment by relying on the exchange of real-time information such as speed, position, and braking status of equipped vehicles with each other, roadside infrastructure, and the Internet [14]. The connected vehicle environment should be leveraged to improve the emergency vehicle’s operations on the roads as “crashes involving emergency vehicles, including ambulances, fire trucks, and police cars, are a substantial problem nationwide”[15] with 368,946 ERV-related crashes reported between 2001 and 2010 [16]. Assistance can be provided to emergency responders on several levels. For instance, automatic crash notification systems are developed to contact the closest emergency center, to quickly evaluate the emergency based on received vehicular data and, hence, to efficiently utilize the required emergency resources [17-20].

ERVs’ operations traveling to and from emergencies can also be enhanced. As congestion increases, their movements are impacted by more crashes and lengthier response times, especially at intersections [21]. Emergency vehicle preemption at signalized intersections is a practice that was proposed in the early 1980s [22] and has been deployed widely in the United States [23]. More recently, an autonomous intersection management system considering the presence of an approaching ERV was proposed [24-26]. In [27, 28], non-ERVs are instructed to maneuver to specific lanes to achieve a split at a critical point in the vehicle queues, allowing ERVs to proceed with maximum speed and minimum disruption through signalized intersections. These studies are limited to intersections and do not mitigate the challenges associated with the ERVs’ movement along transportation links.

Confusion is a serious issue that downstream vehicles face upon hearing the siren and/or visually detecting an ERV. Drivers may not be aware of which lane the ERV wants to use and how to best maneuver to allow the ERV to pass safely and efficiently [29]. Failure to detect the ERV can also result from the
driver’s impaired hearing, inattentiveness, and other background noise [30]. Emergency vehicle warning systems support emergency services as they alert downstream and surrounding vehicles on the presence of a nearby ERV, granting these vehicles more time and relevant information to react efficiently [31-34]. However, these alert systems do not suggest and recommend the best actions to be adopted by downstream vehicles. Even with these messages alerting vehicles of an approaching ERV, vehicles may still fail to respond in a timely manner and to coordinate its movement with the other vehicles adequately, resulting in a slower intra-link ERV path. A study [35] on a two-lane link facilitates the passage of an ERV by requesting the downstream vehicles to shift away from the lower density lane, reserving it for ERV use. The limited road width and assumption that the lower density lane is the best intra-link ERV path constitute limitations as the ERV may desire the use of a specific lane based on its next intersection movement and/or the emergency scene’s location. Another system inducing a lane change maneuver for vehicles obstructing the ERV’s way is proposed in [36]. While a rescue lane for the ERV is freed and improvements in the ERV travel time are observed, this system assumes a predefined ERV path (i.e., it does not generate the best ERV intra-link). Furthermore, the ERV does not receive recommendations about the best maneuvering actions based on the feasible cooperative movements that downstream vehicles can acquire. Similarly, in [37], priority is granted to ERVs on an automated highway system by moving vehicles or platoons of vehicles out of the ERV’s way. The ERV only reacts to the available downstream space that can be provided by the downstream vehicles and its movement may not be optimal. In addition, this study does not accommodate the presence of unconnected (unequipped) vehicles.

The main capabilities of connected vehicles are data availability and exchange. A partial market penetration means omitted information as the unequipped vehicles which are in reality present in the system are not sharing their corresponding data. To evaluate the performance of the connected vehicle applications, mitigating partial market penetration is crucial. The basic idea relies on estimating the positions of the unequipped vehicles using data received by equipped ones [11]. In [38], a microscopic analysis is introduced to dynamically adjust the variable speed limit in a connected vehicle environment. As the model requires the vehicle trajectories of all the vehicles in the system and partial market penetration is possible, a microscopic estimation is introduced in [39]. It consists of approximating the state of unconnected vehicles between two connected vehicles after comparing the actual behaviors, in terms of acceleration and deceleration, of the two connected vehicles with their expected ones generated from a car following model. Goodall et al. [40] also adapted their algorithm to signalized arterials so that unusual observed headways (greater than 14.5 m, i.e., double the space headway of 7.25 m predicted by the Wiedemann model), between connected vehicles stopped in queues, act as indicators of the presence of unconnected vehicles. In [11], a real-time adaptive signal control that is based on a phase allocation algorithm is introduced. Since complete arrival data on all the intersection’s approaches is needed as input, the state estimation of unconnected
vehicles is required. After defining the queuing, slow-down and free-flow regions along the link upstream of the intersection, they developed an algorithm called EVLS (estimation of location and speed) that estimates the unconnected vehicle presence and status in each region. In the queuing region, since the speed is zero, only the presence (position) of an unconnected vehicle is left undetermined and is identified by estimating the queue length. In the slow-down region, the previously discussed estimation technique in [39] is not applicable as the difference between the observed and predicted behaviors in this region may never reach the threshold that triggers vehicle insertions (i.e., vehicles travelling at common states). Subsequently, a new rule is adopted in which an unconnected vehicle is inserted when the observed headway between two consecutive equipped vehicles exceeds the double of the maximum car following distance predicted by the Wiedemann car following model. In free flow conditions, since no interactions among vehicles exit, a crude method is adopted and consists of (1) estimating the total number of vehicles on the link by dividing the number of detected (equipped) vehicles by the penetration rate, (2) uniformly distributing the number of unconnected vehicles (i.e., difference between the total number of vehicles and the number of connected vehicles) among the lanes, (3) randomly assigning the unconnected vehicles at positions along the lane, and (4) setting the speeds of the unconnected vehicles to the observed speed or posted speed limit [11].

In these cases where the connected vehicle applications are expected to return instruction messages to equipped vehicles, it is important to consider that the unequipped vehicles are unable to receive them. According to [41], surrounding traffic conditions is regarded differently by different drivers, under the lack of communication. The sources of information of these unequipped vehicles are limited to mobile apps, global positioning systems, variable message signs or simple observations [42].

The strategies used to account for partial market penetration can also be employed to consider imperfect human compliance and deteriorated vehicle communications. According to [43], “the National Highway Traffic Safety Administration (NHTSA) defined five levels of automation, starting from level 0 which is the basic traditional vehicles with no automation up to level 4 with full self-driving automation”. So, even if a vehicle is equipped with connected vehicles technologies, when dealing with levels of automation from 0 to 3, the degree of human compliance should be taken into account as the driver still has the capability of monitoring and controlling the vehicle. In other words, the driver may disregard the received assistance messages and intervene and resume control at any time in an unpredictable manner. Therefore, a semi-automated connected vehicle (an equipped vehicle with a human driver) that is not complying with the assistance messages can be regarded as an unconnected vehicle. When semi-automation exists, the imperfect degrees of human compliance can be translated to lower market penetration. It is also important to note that vehicle communications are sensitive to external factors such as weather, road topology, and traffic congestion [44]. This is why even in a connected vehicle environment with full market
penetration, due to degrading vehicle communications, communication failures may result and equipped vehicles might lose connectivity, leading to an indirect decrease in the market penetration [11].

In conclusion, when developing new systems and applications based on the connected vehicle technologies, it is important to assess their functionality and performance first by assuming that all vehicles are equipped and second by considering partial market penetration. This helps in avoiding the congestion paradox that occurs when negative impacts arise instead of benefits in terms of congestion and social costs, when compared to the current state [45]. Our work described below accounts for partial market penetration when facilitating the ERV intra-link passage.

4.3. Proposed System

In this paper, Hannoun et al.’s [10] initial approach is extended to optimize the ERV passage over larger distances. The proposed approach relies on the presence of connected vehicle technology first to retrieve the input data about the non-ERVs along the link and second to disseminate the final positions to non-ERVs as well as the instruction messages to the ERV. The approach assumes the presence of a centralized computing server that pre-processes the collected data, runs the ILP, post-processes and stores the ILP output and sends messages to ERV and non-ERVs. The proposed ILP formulation, in its present form, only accommodates a single ERV that increases/decreases its speed based on the maneuvers within the link and the presence of surrounding non-ERVs. The approach is limited to a transportation link with no intersections or non-ERVs entering/leaving the link. The technique used to account for partial market penetrations (discussed further in this section) assumes that the penetration level is known and that the unconnected non-ERVs perceive and follow the behavior of their leader.

4.3.1) System Model

As shown in Figure 4-1, a link is divided into segments, called Initial Ranges (IRs). After collecting data (positions, speeds and deceleration capabilities) about non-ERVs, each IR undergoes an ILP optimization generating the ERV optimal intra-link path and non-ERV positions along a downstream range called the Assignment Range (AR). Next, ERV instruction maneuvers as well as ERV speeds to be adopted are generated incrementally along this AR. The start and end of the corresponding AR along the link is identified based on the non-ERVs initial positions, speeds and deceleration capabilities in each IR (for more details, see Section 4.4.1). Additional preprocessing steps and constraints are introduced to ensure continuity between the ILP optimizations (as shown in Figure 4-2, in the post-processing of the output box).
4. 3. 2) System Description

The roadway is discretized into identical cells of size $L$ by $W$, that is the size of a regular vehicle plus a buffer. A cell is characterized by its $(x, y)$ coordinates, where the $x$ axis denotes the longitudinal motion (i.e., direction of flow) and the $y$ axis refers to the lateral motion (i.e., lane changes). The ERV instructions are generated at every increment, where one increment encompasses a number of cells equal to the ERV longitudinal size plus a buffer of 1. The ERV speeds associated with each increment are expressed in speed stage to take integer values while avoiding impractical acceleration and deceleration rates (see Appendix in [10]). The non-ERVs are labeled by $j$ based on their initial positions in the IR; non-ERVs located on higher $x$ and $y$ coordinates receive larger labels.
It is important to build a connection from one optimization to the other based on the relative position of the current AR with the previous one. The ERV initial lateral position and initial speed while entering the first AR (corresponding to the first IR) is known. For the next ARs, these two input parameters can be deduced from the previous optimization. As shown in Figure 4-1, a gap or an overlap may appear between ARs since each AR’s starting and ending positions are dictated by the minimum final positions of the non-ERVs travelling on its corresponding IR. If a gap exists between ARs (see AR2 and AR3 in Figure 4-1), it is assumed that the ERV maintains a straight path along this gap. For instance, if the ERV exited AR2 on the leftmost lane, it enters AR3 on the leftmost lane as well, with a speed (in stages) that linearly increases starting from the last observed speed on AR1 (the gap implies free adjacent cells allowing for a speed increase up to the ERV maximum allowable speed). In the case of an overlap with the previous AR (see AR1 and AR2 in Figure 4-1), the observed ERV lateral position and speed at the increment in AR 1 that corresponds to the start of AR2 is simply inputted to the optimization of AR2. In addition, in the case of overlap, passing is prohibited between all non-ERVs (in the same and in different IRs) to reduce conflicts. An upper limit is set for the ERV speed to account for the presence of previously assigned non-ERVs on previous ARs. All of these tasks constitute the preprocessing steps that should be performed between each pair of consecutive ARs; they are explained with more details in the next section.

4.3.3) Partial Market Penetration

In the proposed approach, each connected non-ERV receives an instruction message with an optimized location at which it should stop. The unequipped non-ERVs are unable to receive such messages and lack information and assistance about how to react. Each assignment message is vehicle specific which means that variable message signs (VMS) cannot be used. Consequently, it is possible that the unconnected non-ERVs’ driving maneuvers may be inefficient and unsafe [41]. In this study, it is assumed that the unconnected non-ERV drivers are alert and aware of the approaching ERV (use of VMS on the ERV path, when possible) and that these non-ERVs follow the behavior of the non-ERV in front of it and in the same lane regardless if it is equipped or unequipped. For example, unconnected non-ERVs do not overtake the upstream non-ERVs that are braking to reach their stopped positions. This assumption resembles the one used in [46] that consists of adopting a car following model as a control system to monitor the unconnected vehicles.

Unconnected non-ERVs are also unable to share their information. The estimation of their position based on the available information received by the connected vehicles is performed. After thorough examination of the possible techniques to account for unconnected non-ERVs, a hybrid technique, is developed. This approach consists of the following steps:
• Step 1: Determine the possible positions of unconnected non-ERVs based on a distance-based criterion based on Wiedemann car-following model adopted for slow-down region in [11] where a vehicle is likely to be present between two connected vehicles separated by a distance greater than two times the maximum car following distance (SDX). The maximum car following distance (SDX) is computed using the following equation: SDX= AX+EX*BX where: AX is the minimum headway of 21.52 ft, EX is a calibration factor of 1.5 and BX is a calibration factor of $4.5\sqrt{v}$ where v is the minimum of the speed of the vehicle, and the speed of the leader in ft/s. This step identifies the maximum number of unconnected vehicles that can be added based on gaps between connected non-ERVs.

• Step 2: Estimate the number of unconnected non-ERVs (Nu) that should be present on the link based on the number of observed connected non-ERVs and the market penetration level. The total (unconnected and connected) number of non-ERVs is equal to the number of connected (observed) non-ERVs divided by the market penetration level. The number of unconnected non-ERVs is the difference between the total number of non-ERVs and the number of connected non-ERVs. This is the number of unconnected vehicles to be added on the link.

• Step 3: Randomly select Nu positions from the unconnected non-ERV positions determined in Step 1. In fact, Step 1 adds unconnected non-ERVs wherever there is the space to fit one. Hence, the number of unconnected vehicles is overestimated and inputting all these unconnected non-ERV positions results in higher traffic flow than the actual one.

Next, the predicted positions of the unconnected non-ERVs and the observed status of the connected non-ERVs are inputted in the ILP. The ILP with one additional constraint (constraint 43 discussed in Section 4.5.) for induced car-following behavior.

Whichever technique is adopted, the ERV speeds computed using the ILP may have to be adjusted in case the unconnected non-ERV positions are improperly predicted. The ERV may still encounter unconnected non-ERVs that were not detected. The ERV speeds (generated by the ILP) may have accounted for the presence of non-ERVs that do not exist or may have failed to consider present unconnected non-ERVs. Nevertheless, in a semi-automated system, the ERV driver is capable of adapting to these unexpected unconnected non-ERVs and adjusting the ERV speeds accordingly. In the long-term future, a connected vehicle will have an autonomous vehicle’s capability of sensing nearby non-ERVs, hence making the estimation of unconnected non-ERVs more accurate. Yet, in this paper, a system for short-term deployment is proposed without relying on this sensing ability for connected vehicles.
4.4. Preprocessing steps

Prior to each ILP optimization, a set of preprocessing steps are required to prepare the corresponding input data. Some of these steps are solely based on characteristics of the IR undergoing optimization next (discussed in [10]) while other preprocessing steps (introduced here) are basically retrieving output from the last optimization and converting it into input for the next one. Parameter and variable notation are listed in Table 4-1 and Table 4-2, respectively.

Table 4-1: Parameter notation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>Longitudinal size or length of a single cell (in ft)</td>
</tr>
<tr>
<td>$W$</td>
<td>Lateral size of a single (in ft)</td>
</tr>
<tr>
<td>$N$</td>
<td>ERV longitudinal size (in cells)</td>
</tr>
<tr>
<td>$LL$</td>
<td>Number of longitudinal cells in the AR (in cells)</td>
</tr>
<tr>
<td>$Y$</td>
<td>Number of lateral cells in the AR, including traversable shoulders (in cells)</td>
</tr>
<tr>
<td>$J$</td>
<td>Number of non-ERVs in the IR</td>
</tr>
<tr>
<td>$AR_{start}$</td>
<td>Longitudinal index of the start of the AR with respect to the start of its Initial Range</td>
</tr>
<tr>
<td>$t_r$</td>
<td>Non-ERVs reaction time (in seconds)</td>
</tr>
<tr>
<td>$\sigma_j$</td>
<td>Non-ERVs initial speed (in fps)</td>
</tr>
<tr>
<td>$\delta_j$</td>
<td>Non-ERV deceleration rate (in fps$^2$)</td>
</tr>
<tr>
<td>$t_{cc}$</td>
<td>Time of communication and computation (in seconds)</td>
</tr>
<tr>
<td>$x_j'$</td>
<td>Initial longitudinal index of non-ERV $j$ (in the IR)</td>
</tr>
<tr>
<td>$y_j'$</td>
<td>Initial lateral index of non-ERV $j$ (in the IR)</td>
</tr>
<tr>
<td>$b_{jj'}$</td>
<td>Binary parameter equal to 1 when $j$ is behind or to the left of $j'$ ($y_j' \geq y_{j'}$), and 0 otherwise</td>
</tr>
<tr>
<td>$MSD_{j}$</td>
<td>Minimum stopping distance of non-ERV $j$</td>
</tr>
<tr>
<td>$MFP_{j}$</td>
<td>Minimum final position of non-ERV $j$ with respect to the IR start (in cells)</td>
</tr>
<tr>
<td>$x_j''$</td>
<td>Minimum final longitudinal index of vehicle $j$ in cells with respect to the start of the AR</td>
</tr>
<tr>
<td>$c$</td>
<td>Longitudinal FSR cutoff value (in cells)</td>
</tr>
<tr>
<td>$S_{min}$</td>
<td>Minimum allowable ERV speed (in speed stage)</td>
</tr>
<tr>
<td>$S_{free}$</td>
<td>Maximum allowable ERV speed (in speed stage)</td>
</tr>
<tr>
<td>$M$</td>
<td>Large number used to apply the Big M method</td>
</tr>
<tr>
<td>$np^{x,y}$</td>
<td>Binary parameter that takes the value of 1 at cells with longitudinal index $x$ less than or equal to the longitudinal index of the most downstream non-ERV in the previous AR, and 0 otherwise</td>
</tr>
<tr>
<td>$d_{i,j,k}$</td>
<td>Binary parameter that takes the value of 1 when an ERV instruction $k$ should be at increment $i$ and lateral index $y$ imposed and 0 otherwise</td>
</tr>
</tbody>
</table>
 Integer parameter that takes the value of any previously assigned ERV speed environment at the increment $i$ (due to overlap with previous AR), and infinity otherwise.

Final desired ERV lateral position

Binary parameter that takes the value of 0 when a final ERV lateral position is desired and 1 otherwise

Binary parameter that takes the value of 1 if non-ERV $j$ is a connected vehicle and 0 otherwise

Integer parameter referring to the leader of each non-ERV $j$ in the IR

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w^{x,y}$</td>
<td>ERV assignment binary variable equal to 1 if cell $(x,y)$ is included in the ERV path and 0 otherwise (where $x=1,\ldots,LL$ and $y=1,\ldots,Y$)</td>
</tr>
<tr>
<td>$s^i$</td>
<td>ERV speed integer variable referring to the ERV speed at increment $i$ (where $i=1,\ldots,LL/(N+1)$)</td>
</tr>
<tr>
<td>$d^{i,y}_k$</td>
<td>ERV instruction binary variable equal to 1 when instruction $k$ is given to the ERV at increment $i$ and lateral position $y$ (where $i=1,\ldots,LL/(N+1)-1$; $y=1,\ldots,Y$; $k=1$ means move right; $k=2$ means go straight and $k=3$ means move left)</td>
</tr>
<tr>
<td>$v^{x,y}_j$</td>
<td>Non-ERV assignment binary variable equal to 1 if cell $(x,y)$ is allocated to non-ERV $j$ cell and 0 otherwise (where $x=1,\ldots,LL$; $y=1,\ldots,Y$ and $j=1,\ldots,J$)</td>
</tr>
<tr>
<td>$s^i_{env}$</td>
<td>ERV speed environment integer variable referring to the ERV speed at every increment $i$ only based on the ERV surrounding (where $i=2,\ldots,LL/(N+1)$)</td>
</tr>
<tr>
<td>$s^i_{temp}$</td>
<td>ERV temporary speed integer variable referring to the ERV speed at increment $i$ based on the ERV surrounding and previous ERV instruction without being limited to the minimum and maximum allowable ERV speeds (where $i=2,\ldots,LL/(N+1)$)</td>
</tr>
<tr>
<td>$v^i$</td>
<td>Binary variable equal to 1 at increment $i$ if $s^i_{temp} \geq S_{min}$ and 0 otherwise (where $i=2,\ldots,LL/(N+1)$)</td>
</tr>
</tbody>
</table>

Preprocessing steps based on the IR’s characteristics

These preprocessing steps consist of defining, for each IR, the feasible stopping range (FSR) of each non-ERV travelling on it, its corresponding AR’s starting/ending positions and the minimum final longitudinal index of each non-ERV. These steps, introduced in [10], are only based on data from the IR itself (no information needed from previous optimizations). In the first preprocessing step, the computation of the minimum stopping distance (in Equation 1), is slightly modified in this paper to include the distance travelled by the non-ERV during the time for communication and computation which is the time between the data collection and receipt of the instruction (See Section 4. 4. 1) 1) and Appendix B).
1) Identification of the Feasible stopping range of each non-ERV \( j \)

The Feasible Stopping Range (FSR) of each non-ERV \( j \) should be determined before the optimization first to locate the projected assignment (or optimization range) and second to make sure that each non-ERV is assigned to a location in its corresponding FSR (using constraint 6). Based on the reaction time in seconds \( (t^r) \), initial speed in ft/sec \( (\sigma_j) \), deceleration in ft/sec\(^2\) \( (\delta_j) \) and time of communication and computation in seconds \( (t_{cc}) \), the minimum stopping distance \( (MSD_j) \) of each non-ERV \( j \) is determined using Equation (1). Note that \( t_{cc} \) is the time elapsed for data collection, optimization and output dissemination during which the non-ERV is still travelling at its initial speed, this is why the distance travelled during \( t_{cc} \) should be included in the \( MSD_j \) calculations. The \( t_{cc} \) parameter is always the time elapsed between the last data collection and the current time (see Appendix B for additional details). After computing \( MSD_j \) (expressed in number of cells and rounded up to the nearest integer), the minimum final position \( (MFP_j) \) of each non-ERV \( j \) is identified using Equation (2), where \( (x_j') \) is the initial longitudinal position of non-ERV \( j \). The FSR of each non-ERV \( j \) extends \( c \) cells beyond the \( MFP_j \) (i.e., the FSR longitudinal size or FSR cut-off value \( c \) is an input parameter to the ILP and is discussed further in Section 4.4.3).

\[
MSD_j = \text{ceil} \left( \left( t_{cc} \times \sigma_j + t^r \sigma_j + 0.5 \frac{\sigma_j^2}{\delta_j} \right) / L \right) \tag{1}
\]

\[
MFP_j = x_j' + MSD_j \tag{2}
\]

2) Identification of the Assignment Range’s start, end and size

The optimization takes place on the assignment range which is downstream of the IR and includes the FSRs of all non-ERVs in the IR. The AR starts at (ERV size + buffer) cells before the smallest \( (MFP_j) \) (i.e., most upstream FSR) and ends at or after the largest \( (MFP_j + c) \) (i.e., most downstream FSR) to make the AR longitudinal size a multiple of the (ERV size + buffer), to have a discrete number of ERV speeds and ERV instructions.

3) Identification of the minimum final longitudinal index of each non-ERV \( j \)

After defining the start of the AR \( (AR_{\text{start}}) \), the minimum final longitudinal index \( (x_j'') \) of each non-ERV \( j \) is calculated using Equation (3):

\[
x_j'' = MFP_j - AR_{\text{start}} + 1 \tag{3}
\]

4) Market penetration preprocessing steps

The estimation step previously discussed should be implemented prior to each link segment optimization when less than full market penetration exists to predict the position of unconnected non-ERVs along the link. If the ILP is left unchanged (i.e. with no modification specifically done to account for partial market penetrations), it does not distinguish the difference between the connected and unconnected non-
ERVs and may instruct an unconnected non-ERV to stop at a location and to initiate a movement that completely differs from the one that the non-ERV ahead is taking. [For example, an unconnected non-ERV may be required to stop at the leftmost lane while its preceding non-ERV is required to stop at the rightmost lane]. For instance, the ILP fails to consider that the unconnected non-ERVs cannot receive messages with enclosed positions, and that a following behavior needs to be assumed to monitor their actions. Consequently, if the presence (and positions) of unconnected non-ERVs are estimated and input in the ILP, imposing a car-following behavior between a predicted unconnected non-ERV and its preceding non-ERV should be considered. So, a new constraint is added (Equation 43) to the formulation in [10] to impose a car-following behavior by unconnected non-ERVs with respect to their corresponding leader. Two new parameters are introduced: (1) \( t_y^j \) is a binary parameter denoting the type of each non-ERV \( j \). It takes the value of 1 for connected non-ERVs and the value of 0 for unconnected non-ERVs and (2) \( l_e^j \) is an integer parameter referring to the leader of each non-ERV \( j \) in the IR. The leader is the non-ERV ahead and in the same lane (i.e., the non-ERV with the smallest label whose longitudinal position is greater than \( x_j^* \) and lateral position is the same as \( y_j^* \)). An unconnected non-ERV should have its leader in the same IR, since the ILP needs to match the unconnected non-ERV’s movement to the non-ERV ahead. However, a connected non-ERV can have its leader in the next IR; in that case, the non-ERV is the leader of itself \( (l_e^j=j) \).

4. 4. 2) Preprocessing steps converting output of last optimization into input for the next one

These steps are crucial to build continuity from one optimization to the other by retrieving data from the output of the last AR and converting it to input parameters for the next optimization.

1) ERV initial lateral position and speed in the AR

For each IR along the link (except the first IR), the ERV initial lateral position and speed has to be deduced from the previous optimization’s output. In case the next AR (i.e., corresponding to the next IR optimization) overlaps with the previous AR, the ERV initial lateral position and speed generated at the increment in the previous AR that coincides with the first increment of the next AR is retrieved. If a gap exists between the next and previous ARs, the ERV initial lateral position and speed are deduced by assuming that the ERV maintains a straight path after exiting the previous AR and increases its speed linearly up to \( S_{\text{free}} \).

2) Non-ERV positions in the previous AR

The binary parameter \( (np^{x,y}) \) at each cell in the upcoming optimization should be determined. This parameter is introduced and used in constraint (39) to ensure that passing is prohibited among non-ERVs in different ARs. It takes the value of 1 at and before the position of the most downstream vehicles in the
previous AR and 0 otherwise. Hence, when no overlap with the previous AR exists or when optimizing the very first IR along the link, the \((np^{x,y})\) at all cells takes the value of 0.

3) **ERV instructions in the previous AR imposed**

The binary parameter \((d_{t_k^{i,y}})\) takes the value of 1 if an ERV instruction \(k\) was generated at increment \(i\) and lateral index \(y\) in the previous optimization. When an overlap with the previous AR exists, some increments belong to more than one ARs at the same time. So, the ERV instructions should be consistent. Different ERV instructions cannot be generated at the same increment. This is why, ERV instructions previously generated over the AR overlap and delineating the ERV path up to the most downstream non-ERV position on the previous AR is fixed and previously set for the upcoming optimization (see Appendix C). When no overlap with the previous AR exists, the \((d_{t_k^{i,y}})\) parameters take the value of 0.

4) **ERV speeds in the previous AR**

A final parameter \((su^i)\) at increment \(i\) in the next AR should be identified. This integer parameter is an ERV speed limit. As previously discussed, when the next AR overlaps with the previous one, the ERV initial speed of the next AR is retrieved from the previous output. Yet, the ERV speed at each increment in the next AR should take into consideration the presence of non-ERVs along the overlap. The \((su^i)\) parameter at increment \(i\) along the overlap takes the value of the speed environment obtained at the corresponding increment in the previous AR (the speed environment in the speed that is only limited by the number of surrounding non-ERVs). When the overlap ends or in cases when overlap does not exist, the \((su^i)\) parameter takes the value of infinity.

4. 4. 3) Dynamic FSR cut-off input value \(c\)

Each non-ERV is allowed to stop within its corresponding FSR which starts at its minimum final position \(MFP\) (computed based on its minimum stopping distance \(MSD\)) and which extends \(c\) cells beyond the \(MFP\) (i.e., the length of the FSR is \(c\)). The higher the \(c\) value, the larger the problem size, as each non-ERV has more final position alternatives and the total AR’s longitudinal size (optimization range) becomes higher by default. In cases when the ILP fails to generate an optimal solution using a given \(c\) value, re-running the ILP with a higher \(c\) value may help find a solution as non-ERVs have more space along which to spread. The system increases the \(c\) value incrementally by 1 until an optimal solution is found. The FSR cut-off initial \(c\) is the value that is used to run the ILP for the first time. Using a static initial \(c\) means assigning the same initial \(c\) value to all IRs, regardless of each IR’s characteristics; then, in cases when infeasibility arises after trying to solve the ILP for a given IR, the \(c\) value of this IR is increased incrementally until a solution is found. It is computationally intensive to continue trying to solve the ILP while incrementally increasing \(c\), this is why it was motivating to develop a preprocessing step that dynamically (i.e., depending on scenario-specific parameters) finds the minimum initial \(c\) value that should
be used in the first IR optimization and that potentially leads to an optimal solution faster. Three rules of thumb are developed to determine the dynamic (required) FSR cut-off initial value $c$.

1) **Rule 1: FSR and no passing constraint**

If $j$ is travelling at a higher speed than $j'$, the FSR of $j$ may start after the end of the FSR of $j'$ (as shown in Figure 4-3). Since passing among vehicles is prohibited, no feasible solution is found in that case. Yet, if $c$ is increased until the FSR of $j'$ overlaps with the FSR of $j$, an optimal solution can be generated.

For each pair of non-ERVs $j$ and $j'$ such as $j'$ is downstream of $j$:

```plaintext
if (endFSR[j'] < startFSR[j]) { // if the FSR of $j'$ ends before the start of the FSR of $j$
  c = startFSR[j] - startFSR[j']; // the required $c$ value
}
```

Select the maximum $c$; // the maximum $c$ value among all pairs is selected

![Figure 4-3: Non-ERVs $j$ and $j'$ with corresponding FSR (when speed of $j$ is higher than the one of $j'$)](image)

For simplicity, the ILP is executed on an IR with only 2 non-ERVs with speeds that force the FSR of a non-ERV $j$ to be located downstream of the FSR of non-ERV $j'$ (where $j'$ is initially a vehicle downstream), as shown in Figure 4-3. The ILP is run first without the integration of the new rule and with an initial $c$ value equal to 2. The ILP did not generate a feasible solution for this initial $c$ value, so the code increased the $c$ until reaching the value 7 that led to a feasible solution. In that case, ILP computation time turned out to be 0.987 second (with time lost to iterate from a $c$ value of 2 to 7). Instead, the ILP is executed with an initial $c$ value of 7, identified using rule 1. This dynamic identification of the minimum required $c$ value led to a decrease in the computation time that turned out to be 0.401 second (no time is spent on iterations). Similar tests were applied on an IR with 3 non-ERVs, and the decrease in computation times due to the addition of this new rule was even more significant (results shown in Table 4-3).

**Table 4-3: Computation times with and without rule 1**

<table>
<thead>
<tr>
<th>Computation time in seconds</th>
<th>Without rule 1</th>
<th>With rule 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 vehicles</td>
<td>0.987</td>
<td>0.401</td>
</tr>
<tr>
<td>3 vehicles</td>
<td>1.454</td>
<td>0.453</td>
</tr>
</tbody>
</table>
2) Rule 2: available space vs required space

When the available number of cells for the ERV movement is less than that required, no solution can be found; hence the increase in the c value is needed. This mostly occurs on highly congested link segments. Increasing the c value increases the AR longitudinal size, thus allowing the non-ERVs to take more dispersed positions and to free a path for the ERV. The available number of cells for the ERV movement is equal to the number of cells in the AR where non-ERVs can stop minus the number of non-ERVs. The number of cells in the AR where non-ERVs can stop is the total number of cells in the AR minus (1) the number of cells before the smallest $MFP_j$ or the number of cells in the overlap with the previous AR, whichever is larger and (2) the number of cells added at the end of the AR to make the AR longitudinal size a multiple of the ERV size + buffer. Due to the no passing rule across ARs, non-ERVs in the ith AR can only stop after the last non-ERV in the (i-1)th AR. This is why the available space does not include the cells in the overlap between ARs. As for the required number of cells for the ERV movement, it is equal to the length of the AR part where the non-ERVs can stop and includes additional cells required to make (LC) lane changes. The ERV is assigned to one cell at each x index; yet, when a lane change is initiated, up to 2 additional cells are freed for the ERV. The minimum c value should give the ERV the freedom to perform up to 2 LC lane changes. Lane changes may result in better ERV speeds, so the ERV should not be forced to take a straight path just because there is only space for a straight movement. The improvements observed in terms of the computation times are shown after the discussion of the third rule.

3) Rule 3: FSR and AR overlap

Due to the no passing rule across ARs, non-ERVs in the i^{th} AR can only stop after the last vehicle in the (i-1)^{th} AR. Infeasibility arises when the FSR of a non-ERV (v1 in AR 2) ends at or before the last assigned vehicle in the previous AR (v2 in AR 1), as shown in Figure 4-4. So, the c in AR 2 needs to be incrementally increased (by 2 in that case) until a feasible solution is found (i.e., so that v1 in AR 2 can stop after v2 in AR 1).

For each non-ERV j:

```
if (endFSR[j] <= x index of the most downstream non-ERV in previous AR){
    c = x index of the most downstream non-ERV in previous AR - startFSR[j] +1;
}
```

Select the maximum c;
According to Table 4-4, if a static initial \( c \) value is adopted (no rules adopted), for the optimization of IR 2 (i.e., AR 2) with 15 non-ERVs, no optimal solution is obtained with a \( c \) equal to 2 so it had to increase from 2 to 3 and the computation time was 0.671 second. If the preprocessing rules are used with LC=0, the initial required \( c \) value was identified (equal to 3) and inputted before solving the ILP, so the computation time decreased to 0.37 second for AR 2. If the preprocessing rules are considered with LC equal to 2, the obtained initial \( c \) value is 5. It is larger because the minimum number of cells required for the ERV movement with 2 lane changes is higher. The computation time for AR 2 is 0.654 second; hence no notable computation time improvement is noted due to the increase in problem size. Similar trends were observed in the scenarios with 24 non-ERVs in IR 2. Selecting the LC value is subjective. If the ERV is initially positioned on an edge lane and is not required to exit the AR from a specific lane, then an LC of 0 is reasonable since, based on [10], the ERV is capable of reaching higher ERV speeds when traveling on the rightmost or leftmost lanes. However, if the ERV is on a middle lane, and if the road is relatively congested (i.e., freeing both sides on the ERV path at all increments is not possible), then it is important to increase the LC value because a lane change may help in achieving higher speeds. In that case, the LC value can be the number of lane changes required to move from the initial lateral position to the closest edge when no final lateral ERV position is imposed; otherwise it can be equal to the number of lane changes required to move from the initial lateral position to the final imposed lateral position.
Table 4: Computation times (with rules vs without rules), where the initially inputted c value is 2 cells when no rules in used

<table>
<thead>
<tr>
<th>Without rules</th>
<th>With rules 1,2 and 3 (LC=0 lane change)</th>
<th>With rules 1,2 and 3 (LC=2 lane changes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computation time</td>
<td>c</td>
<td>Computation time</td>
</tr>
<tr>
<td>15 vehicles in IR 2</td>
<td>AR1: 0.345 AR2: 0.671</td>
<td>3</td>
</tr>
<tr>
<td>24 vehicles in IR 2</td>
<td>AR1: 0.35 AR2: 1.609</td>
<td>4</td>
</tr>
</tbody>
</table>

In conclusion, determining a dynamic FSR cut-off initial c value using simple computations based on three rules of thumb help in reducing computation time spent during unnecessary iterations to reach optimality. However, these rules do not identify, under all scenarios, the c value that leads to an optimal solution. Iterations may still be required even after determining the dynamic FSR cut-off initial c value based on scenario-specific parameters.

4.5. ILP formulation

The objective function of each IR optimization maximizes the ERV’s speeds (\(s^i\)) at each increment (first component in Equation 4), maximizes the number of free cells adjacent to the ERV path through the maximization of (\(s^i_{env}\)) at each increment (second component in Equation 4), and minimizes the longitudinal indices of the final non-ERV positions (third component in Equation 4). There may be alternate optima for each scenario. Alternative optimal solutions may have different non-ERV positions that lead to the same ERV path and ERV speeds. Since passing among non-ERVs is prohibited (Equations 8 and 39), the alternative optimal solutions with the more upstream final non-ERV positions is preferred and selected. This way the non-ERVs of the following IR optimization are allowed to stop on cells with smaller longitudinal indices, hence more efficiently utilizing the downstream space. The summation of (\(v_j^{x,y}x\)) is multiplied by a very small factor, so that the third component favors one alternative solution (the one with the smallest sum of non-ERV longitudinal positions) without having an impact on the selection of the optimal ERV path and ERV speeds. Equal weights are assigned for the first and second components (\(\alpha_1\)) and (\(\alpha_2\)) respectively as this is the most unbiased combination based on the weight analysis performed in [10].

\[
\text{Maximize } z = \alpha_1 \sum_{i=2}^{LL/(N+1)} s^i + \alpha_2 \sum_{i=2}^{LL/(N+1)} s^i_{env} - 0.00001 \sum_x \left( \sum_y \sum_j v_j^{x,y}x \right)
\]  

(4)
The same constraints that appeared in [10] were employed and are listed below (for more details refer to [10]):

- Each cell can be occupied by only one vehicle (ERV or non-ERV).
  \[ w^x,y + \sum_{j=1}^{j} v^x,y_j \leq 1; \; \forall (x,y) \quad (5) \]

- Each non-ERV is allocated to one cell in its FSR and to exactly one cell in the AR.
  \[ \sum_{x=x_j', y=1}^{x+c} v^x,y_j = 1; \; \forall j \quad (6) \]
  \[ \sum_{x=1, y=1}^{LL, y} v^x,y_j = 1; \; \forall j \quad (7) \]

- Passing among non-ERVs in the same AR is prohibited. The Big M method (where M is a large number) is used for the linearization of this constraint.
  \[ \sum_{x=1}^{x-1} \sum_{y=1}^{y} v^x,y_j \leq M(1 - \sum_{y=1}^{y} v^x,y_j); \; \forall x \geq 2; \forall j' > j \quad (8) \]

- Weaving among non-ERVs is prohibited.
  \[ \sum_{y'=1}^{y-1} \sum_{x=1}^{LL} v^x,y'_{j'} = M \left( |x_j' - x_j| \right) + M \left( 1 - \sum_{x=1}^{LL} v^x,y_j \right); \; \forall y; \forall j' > j \quad (9) \]
  \[ (1 - b_{jj'}) \times \left( \sum_{y'=1}^{y-1} v^x,y'_{j'} \right) \times \left( |y_j' - y_j| \right) \leq M (1 - v^x,y_j); \; \forall x; \forall y; \forall j' > j \quad (10) \]
  \[ b_{jj'} \times \left( \sum_{y'=y+1}^{y} v^x,y'_{j'} \right) \times \left( |y_j' - y_j| \right) \leq M (1 - v^x,y_j); \; \forall x; \forall y; \forall j' > j \quad (11) \]

- A passing lane that consists of one empty cell at every x is freed for the ERV.
  \[ \sum_{y=1}^{y} \sum_{j=1}^{j} v^x,y_j \leq Y - 1; \; \forall x \quad (12) \]

- Only one ERV instruction is generated at each increment.
  \[ \sum_{y=1}^{y} \sum_{k=1}^{3} d^i,y_k = 1; \; \forall i = 1, \ldots, LL/(N + 1) - 1 \quad (13) \]

- No right/left lane change is allowed when the ERV is on the rightmost/leftmost lane.
  \[ d^i,1 = 0; \; \forall i \quad (14) \]
\begin{equation}
d_{2}^{i,y} = 0; \quad \forall i \quad (15)
\end{equation}

- ERV assignment and ERV instruction variables has to be linked to ensure continuous longitudinal and lateral motions.

\begin{equation}
\sum_{y=1}^{Y} w^{x,y} = 1; \quad \forall x \quad (16)
\end{equation}

\begin{equation}
d_{1}^{i,y} \leq \frac{w^{i(N+1),y} + w^{i(N+1)+t,y-1}}{2}; \quad \forall i = 1, ..., LL/(N + 1) - 1; \quad \forall t = 1,2, ..., N + 1; \quad \forall y > 1 \quad (17)
\end{equation}

\begin{equation}
d_{1}^{i,y} \geq w^{i(N+1),y} + w^{i(N+1)+t,y-1} - 1; \quad \forall i = 1, ..., LL/(N + 1) - 1; \quad \forall t
\end{equation}

\begin{equation}
= 1,2, ..., N + 1; \quad \forall y > 1 \quad (18)
\end{equation}

\begin{equation}
d_{2}^{i,y} \leq \frac{w^{i(N+1),y} + w^{i(N+1)+t,y}}{2}; \quad \forall i = 1, ..., LL/(N + 1) - 1; \quad \forall t
\end{equation}

\begin{equation}
= 1,2, ..., N + 1; \quad \forall y \quad (19)
\end{equation}

\begin{equation}
d_{2}^{i,y} \geq w^{i(N+1),y} + w^{i(N+1)+t,y - 1} - 1; \quad \forall i = 1, ..., LL/(N + 1) - 1; \quad \forall t
\end{equation}

\begin{equation}
= 1,2, ..., N + 1; \quad \forall y \quad (20)
\end{equation}

\begin{equation}
d_{3}^{i,y} \leq \frac{w^{i(N+1),y} + w^{i(N+1)+t,y+1}}{2}; \quad \forall i = 1, ..., LL/(N + 1) - 1; \quad \forall t
\end{equation}

\begin{equation}
= 1,2, ..., N + 1; \quad \forall y < Y \quad (21)
\end{equation}

\begin{equation}
d_{3}^{i,y} \geq w^{i(N+1),y} + w^{i(N+1)+t,y+1} - 1; \quad \forall i = 1, ..., LL/(N + 1) - 1; \quad \forall t
\end{equation}

\begin{equation}
= 1,2, ..., N + 1; \quad \forall y < Y \quad (22)
\end{equation}

- The cells constituting the ERV path should not be occupied by non-ERVs. It is assumed that the ERV needs N+1 downstream cells empty in the same y open when going straight. However, when performing a lane change, it is assumed that the ERV needs N downstream cells in the same y to maneuver towards the future lane in which N+1 downstream cells are emptied.

\begin{equation}
\sum_{x=t'}^{t''} \sum_{j=1}^{L} v_{j}^{x,y} \leq M(1 - d_{2}^{i,y}); \quad (23)
\end{equation}

\begin{equation}
\sum_{x=t'}^{t''} \sum_{j=1}^{L} v_{j}^{x,y} + \sum_{x=t'}^{t''} \sum_{j=1}^{L} v_{j}^{y-1} \leq M(1 - d_{1}^{i,y}); \quad (24)
\end{equation}

\begin{align*}
\forall y > 1; \quad \forall i = 1, ..., \frac{LL}{N + 1} - 1; \quad t' = (N + 1)i + 1; \quad t'' = (N + 1)i + (N + 1); \quad T'' = (N + 1)i + N
\end{align*}
\[ \sum_{x=t}^{T'} \sum_{j=1}^{f} v_j^{x,y} + \sum_{x=t'}^{T'} \sum_{j=1}^{f} v_j^{x,y+1} \leq M(1 - d_3^{i,y}); \quad (25) \]

\[ \forall y < Y; \forall i = 1, \ldots, LL/(N + 1) - 1; t' = (N + 1)i + 1; T' = (N + 1)i + (N + 1); T'' = (N + 1)i + N \]

- The ERV speed environment \( s_{env}^{i+1} \) is constrained by the number of nearby stopped vehicles around its next movement. The ERV speed environment variables, that only take into consideration the surrounding, are limited by an upper limit and defined separately to be maximized in the objective function.

\[ s_{env}^{i+1} \leq s^i + 1 - \sum_{j=1}^{f} v_j^{(N+1)i+t-1} - \sum_{j=1}^{f} v_j^{(N+1)i+t-1} + M(1 - w^{(N+1)i+t}); \quad (26) \]

\[ \forall i = 1, \ldots, LL/(N + 1) - 2; \forall t = 0,1, \ldots, N + 2; \forall y = 2,3, \ldots, Y - 1 \]

\[ s_{env}^{LL/(N+1)} \leq s_{env}^{LL/(N+1)-1} + 1 - \sum_{j=1}^{f} v_j^{LL-(N+1)+t-1} - \sum_{j=1}^{f} v_j^{LL-(N+1)+t-1} + M(1 - w^{LL-(N+1)+t}); \quad (27) \]

\[ \forall t = 0,1, \ldots, N + 1; \forall y = 2,3, \ldots, Y - 1 \]

\[ s_{env}^{i+1} \leq s^i + 1 - \sum_{j=1}^{f} v_j^{(N+1)i+t+1} + M(1 - w^{(N+1)i+t+1}); \forall i = 1, \ldots, LL/(N + 1) - 2; \forall t \]

\[ = 0,1, \ldots, N + 2 \quad (28) \]

\[ s_{env}^{LL/(N+1)} \leq s_{env}^{LL/(N+1)-1} + 1 - \sum_{j=1}^{f} v_j^{LL-(N+1)+t+1} + M(1 - w^{LL-(N+1)+t+1}); \forall t \]

\[ = 0,1, \ldots, N + 1 \quad (29) \]

\[ s_{env}^{i+1} \leq s^i + 1 - \sum_{j=1}^{f} v_j^{(N+1)i+t-1} + M(1 - w^{(N+1)i+t-1}); \forall i \]

\[ = 1, \ldots, LL/(N + 1) - 2; \forall t = 0,1, \ldots, N + 2 \quad (30) \]

\[ s_{env}^{LL/(N+1)} \leq s_{env}^{LL/(N+1)-1} + 1 - \sum_{j=1}^{f} v_j^{LL-(N+1)+t-1} + M(1 - w^{LL-(N+1)+t-1}); \forall t \]

\[ = 0,1, \ldots, N + 1 \quad (31) \]

- The ERV temporary speed variables \( s_{temp}^i \) at each increment take into account the surrounding conditions (as the ERV speed environment variables) as well as the ERV instruction given at the previous increment. Hence, the upper limit for \( s_{temp}^i \) at increment \( i \) is set to be the ERV speed
environment variable at that same increment. Next, it is assumed that the ERV has to decrease its speed by one after performing a lane change while it can increase its speed by 1 if the ERV moved forward.

\[ s_{temp}^{i+1} \leq s_{env}^{i+1}, \forall i = 1, ..., LL/(N + 1) - 1 \]  
\[ s_{temp}^{i+1} \leq s^i + 2 \sum_{y=1}^{y} d_{z}^{i,y} - 1; \forall i = 1, ..., LL/(N + 1) - 1 \]

- The ERV speed \((s_{temp}^{i+1})\) cannot increase beyond the maximum allowable speed \(S_{free}\).

\[ s^{i+1} \leq S_{free}, \forall i = 1, ..., LL/(N + 1) - 1 \]  
- The ERV speed \((s_{temp}^{i+1})\) cannot decrease below the minimum allowable speed \(S_{min}\). This is why the ERV speed should be equal to the maximum of \((s_{temp}^{i+1})\) (i.e., minimum of \(s_{env}^{i+1}\) and \(s^i + \text{instruction factor} \)) and \(S_{min}\).

\[ s^{i+1} \geq S_{min}, \forall i = 1, ..., LL/(N + 1) - 1 \]  
\[ s^{i+1} \geq s_{temp}^{i+1}, \forall i = 1, ..., LL/(N + 1) - 1 \]  
\[ s^{i+1} \leq M(1 - v^{i+1}) + s_{temp}^{i+1}, \forall i = 1, ..., LL/(N + 1) - 1 \]  
\[ s^{i+1} \leq Mv^{i+1} + S_{min}, \forall i = 1, ..., LL/(N + 1) - 1 \]

New constraints (39-43) have emerged with the development of the sequential ILP optimization approach.

- Passing among vehicles in different ARs is prohibited by Equation (39) (non-ERVs in the current optimization can only stop after the most downstream non-ERV position of the previous AR). The \((np^{x,y})\) is a parameter that takes the value of 1 when the longitudinal index \(x\) of the cell \((x, y)\) is less than or equal to the most downstream non-ERV longitudinal position in the previous AR and 0 otherwise.

\[ \sum_{i}^{j} v_{j}^{x,y} + np^{x,y} \leq 1; \forall (x, y) \]

- In the case of overlapping ARs, the ERV instructions generated from the previous AR at the same increment \(i\) and occurring before the most downstream non-ERV position in the previous AR, should be imposed. Subsequently, in case of AR overlap, the previously obtained ERV path up to the most downstream non-ERV in the previous AR is maintained and the remainder of the path is subject to change as it navigates a link section where new non-ERVs from the current AR will stop (for further details refer to Appendix C). The \((d_{x}^{i,y})\) is a parameter that takes the value of 1 if instruction \(k\) should be imposed at increment \(i\) and lateral index \(y\), and 0 otherwise.

\[ d_{x}^{i,y} \geq d_{x}^{i,y}; \forall k; \forall i = 1, ..., LL/(N + 1) - 1; \forall y \]

- The ERV speed environment at increment \(i\) is limited by any previously obtained speed environment at increment \(i\) to take into consideration the presence of non-ERVs in the previous AR in the case of
overlap. A parameter \((s_u^{i+1})\) takes the value of the \(s_{env}^{i+1}\) obtained at the same increment \(i\) from the previous AR (if existing) and infinity otherwise.

\[
s_{env}^{i+1} \leq s_u^{i+1}; \quad \forall i = 1, ..., LL/(N + 1) - 1 \tag{41}
\]

- In many cases, the ERV has to exit the AR from a specific lateral position. A new constraint is added to impose the desired last ERV lateral position \((y_f)\) when needed. The parameter \((end)\) takes the value of 0 when a final ERV lateral position is wanted and 1 otherwise.

\[
w^{(LL,y_f)} + end \geq 1 \tag{42}
\]

- The technique adopted to account for partial market penetrations required the introduction of a constraint that implies following behavior by unconnected non-ERVs. Equation (43) ensures that: if the leader of an unconnected non-ERV \(j\) stops at cell \((x,y)\) (i.e., \(t_y = 0\) and \(v_{x,y}^{x,y} = 1\)), then, the non-ERV \(j\) has to stop in the same lane \(y\) (i.e., \(\sum_{x'=1}^{x-1} v_j^{x',y} = 1\)) and not necessarily in the cell behind it to avoid passing among vehicles. This equation ensures that only unconnected non-ERVs follow their respective leader regardless of the type of the latter. In the case of full market penetration, all non-ERVs have \(t_y = 1\), hence this constraint is not binding.

\[
\sum_{x'=1}^{x-1} v_j^{x',y} + t_y \geq v_{x,y}^{x,y}; \quad \forall j; \forall x = 2..LL, \forall y \tag{43}
\]

- The initial ERV lateral position and speed at the start of the AR are determined based on the preprocessing steps and corresponding constraints are included accordingly.

4. 6. Experimental analysis

In this section, the sequential optimization is examined further to provide insights and recommendations that are useful when applying this sequential optimization approach. In [10], a sensitivity analysis was conducted to study the effects of the initial parameters appearing in a single link segment optimization on the computation time and ERV path and speeds. The following parameters were evaluated: road type, road composition, ERV size, ERV initial position, ERV initial speed, number of non-ERVs, non-ERVs’ initial positions, and non-ERVs’ speeds. Higher computation times are observed on narrower roads, with larger numbers of non-ERVs, and when their initial positions are clustered.

In this paper, the focus lies in evaluating new considerations for the sequential optimization: (1) the number/size of IRs in a link and (2) the grouping of vehicles in IRs. The IR number and IR size are two indirectly proportional parameters; as the size of the IR increases the number of IRs decreases for a predefined link length. Hence, both parameters are evaluated in a single test (Test A). Furthermore, the components that characterize an efficient grouping of non-ERVs in IRs are evaluated in Test B.
In addition to the sequential optimization factors in Tests A and B, in Test C, a sensitivity analysis is performed to evaluate the performance of the system when a portion of the downstream non-ERVs is unconnected. The performance of the model (in which the ILP is executed with the estimated unconnected non-ERVs positions) is investigated by assessing the implications engendered after replacing the set of estimated unconnected non-ERVs by the set of actual unconnected vehicles. Market penetrations ranging from 70% to 100% are tested for increasing v/c ratios of 0.75, 0.85 and 0.95.

In Test D, for each v/c ratio, the output is compared to the local practice “go to the nearest edge” and benefits in terms of ERV travel times are computed.

The sequential optimization is coded in java with the AMPL API. The ILP is solved using the CPLEX solver which relies on the branch and bound technique. The java code is executed on a MacOS machine with a 3.1 GHz Intel Core i5 processor and 8 GB 2133 MHz LPDDR3 memory.

4. 6. 1) Test A: IR size/number

The approach in [10] could not find solutions within a reasonable amount of time for longer segments, so, the approach presented in this paper divides the link into smaller link segments (called IRs) that undergo an ILP optimization sequentially. This test consists of varying the number of IRs in the same link of 1575 ft-length (equivalent to 75 longitudinal cells) for three v/c ratios (0.75, 0.85, and 0.95). The ILP parameters are shown in Table 4- 5. Although the homogeneity of IR size is not mandatory in the approach, the sizes of the IRs in this test are the same and increase/decrease uniformly. As the IR size decreases, the problem size of a single IR drops because the number of decision variables and constraints decline. So, the time duration elapsed between the data collection and the instruction receipt is reduced, hence strengthening our assumptions (about no passing and no weaving among vehicles during this time duration and about the estimation/behavior of unconnected vehicles discussed further in Section 4. 6. 3)). According to Figure 4-5, as expected, improvements in computation times are obtained as the link is divided into smaller IRs (i.e., a higher number of ILP optimizations with the link). The higher the v/c ratio, the more significant the average computation time’s improvement. For instance, an average computation time’s decrease from approximately 8.31 seconds to 0.33 seconds for a v/c ratio of 0.95 is observed compared to a decrease from 2.81 seconds to 0.27 seconds for a v/c ratio of 0.75. For all v/c ratios, switching from 10 IRs to 15 IRs did not lead to noticeable computation time decreases (0.07 second for v/c ratio of 0.75, 0.11 second for v/c ratio of 0.85 and 0.14 seconds for v/c ratio of 0.95), meaning that further reduction of IR sizes is not worthwhile. Decreasing the number of IRs does not guarantee benefits as it may jeopardize the optimized ERV path. Determining the optimal ERV path on consecutive short link segments may not be the same as the global optimal ERV path on larger link segments. In our tested scenarios, the decrease in IR size did not generate different ERV path/ERV speeds. Nevertheless, even if the decrease of IR sizes does not impact the ERV path, one should select the minimum IR size beyond which no improvement in computation times
are observed. Smaller IR sizes mean a larger number of IRs/ILP optimizations (within a defined link) and this, in return, translates into higher volumes of data exchange over the network that potentially lead to more communication failures and packet losses. In addition, shorter IRs limit the applicability of the market penetration approach that is discussed further in Section 4.6.3).

Table 4-5: Parameters used for scenarios in Test A

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Notation</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell size</td>
<td>$L$</td>
<td>21 feet</td>
</tr>
<tr>
<td>ERV longitudinal size</td>
<td>$N$</td>
<td>2 cells</td>
</tr>
<tr>
<td>Link lateral size</td>
<td>$Y$</td>
<td>4 (no shoulder)</td>
</tr>
<tr>
<td>Reaction time</td>
<td>$t_r$</td>
<td>2.5 seconds</td>
</tr>
<tr>
<td>Non-ERV deceleration rate</td>
<td>$dec$</td>
<td>11.2 ft/sec$^2$</td>
</tr>
<tr>
<td>ERV maximum speed</td>
<td>$S_f$</td>
<td>Stage 8</td>
</tr>
<tr>
<td>ERV minimum speed</td>
<td>$S_{min}$</td>
<td>Stage 1</td>
</tr>
<tr>
<td>ERV initial position</td>
<td>$y_i$</td>
<td>3</td>
</tr>
<tr>
<td>ERV initial speed</td>
<td>$S_i$</td>
<td>4</td>
</tr>
<tr>
<td>FST cut-off initial value (static)</td>
<td>$c$</td>
<td>3</td>
</tr>
<tr>
<td>Total number of non-ERVs on the link</td>
<td>$J$</td>
<td>$v/c=0.75: J=72$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$v/c=0.85: J=100$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$v/c=0.95: J=128$</td>
</tr>
</tbody>
</table>

Figure 4-5: Average computation time per IR for different numbers of IRs within the link and different levels of congestion

4.6.2) Test B: Non-ERVs’ grouping in IR

This test consists of varying the grouping of non-ERVs while maintaining the same number of IRs (in our case 5 IRs) in the link of 1575 ft length. The goal of this test is to demonstrate the implications that different grouping of vehicles in IRs have on the computation times. As illustrated in Figure 4-6, the top
grouping pattern excludes the first gap of 210 ft between IR 1 and IR 2 and the gap of 105 ft between IR 2 and IR 3, unlike the bottom grouping pattern in which the first gap is part of IR 1 and the second gap is part of IR 2. The computation times of each of the ARs in each grouping pattern are shown in Table 4-6. Results show how the computation times increased by nearly 50% for the first and second IRs and this is mainly due to the additional empty cells included in each IR. Including these large gaps in IRs added to the computation times but did not affect the optimized (generated) ERV path (i.e., same ERV travel time of 32.66 seconds). Regarding IR 3, IR 4, and IR 5, slight variations are observed due to the variation of the IR size. Consequently, this test acts as a confirmation that to efficiently group non-ERVs in IRs, the insertion of large gaps should be avoided. In other words, large gaps can act as the delimiters between IRs, and thus should not be included in any optimization.

![Figure 4-6: Groupings of non-ERVs in IRs](image)

**Table 4-6: Average computation time per AR per grouping shown in Figure 4-4**

<table>
<thead>
<tr>
<th>Average computation time (in seconds)</th>
<th>Top grouping</th>
<th>Bottom grouping</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1</td>
<td>0.571</td>
<td>0.851</td>
</tr>
<tr>
<td>AR2</td>
<td>0.842</td>
<td>1.277</td>
</tr>
<tr>
<td>AR3</td>
<td>1.827</td>
<td>1.671</td>
</tr>
<tr>
<td>AR4</td>
<td>2.166</td>
<td>2.273</td>
</tr>
<tr>
<td>AR5</td>
<td>1.928</td>
<td>1.636</td>
</tr>
</tbody>
</table>

4. 6. 3) Test C: Market penetration tests

According to [10], when an ERV is initially on the rightmost lane, more benefits in terms of ERV speeds were recorded. Hence, the scenarios in the following tests have the ERV entering from the rightmost lane. Three levels of congestion (v/c ratios) are evaluated: 0.75, 0.85 and 0.95. For each level of congestion, market penetration levels ranging from 70% to 100%, (at 10% increments) are tested. Based on the positions of the connected non-ERVs, the presence of unconnected non-ERVs is estimated using a distance criterion (as previously discussed in Section 4. 3. 3)).
Next, the implications that the presence of the actual, unconnected non-ERVs have on the generated ERV movement and the connected non-ERVs following their corresponding instructed message is investigated. With the estimated, unconnected non-ERV positions, the ILP is executed and the optimal ERV path and connected non-ERV final positions are recorded. Next, the ILP is re-executed with the actual, unconnected non-ERV data (instead of the estimated, unconnected non-ERV data) after imposing the previously obtained ERV path and the connected non-ERV relaxed final positions (through the addition of new constraints). Instead of assigning the connected non-ERV exactly at their instructed final positions generated using the ILP with the estimated, unconnected non-ERVs’ positions, each connected non-ERV is allowed to adjust its final position by one extra cell in the longitudinal direction depending on the situation. The connected non-ERV’s driver is assumed to be well-aware of his/her surroundings and reacts in a way to accommodate any unconnected non-ERV that is blocking the ERV’s path. This test evaluates whether it is feasible for the actual, unconnected non-ERVs to safely stop at a cell that is not utilized by any connected non-ERV without jeopardizing the no passing/ no weaving rule and without affecting the ERV optimal path and while following the behavior of their leader. It is an indirect assessment of whether our proposed model is acting as an efficient and optimal reservation approach with partial market penetration. Figure 4-7 shows the percentage of connected non-ERVs that had to adjust their longitudinal position by one cell to accommodate the presence of the actual, unconnected vehicles without impacting the ERV intra-link path and speed and while ensuring that the actual, unconnected vehicles are following the behavior of their leader and not weaving nor passing others. It is important to note that the required adjustment is a distance of only one cell, which is as relatively small (21 feet only). At higher v/c ratios, the vehicles are closer to each other, thus, when one connected non-ERV adjusts its position by one cell to accommodate an actual, unconnected non-ERV, many other connected non-ERVs are forced to do the same. In less dense scenarios, vehicles’ final positions are more spaced; hence if a connected non-ERV adjusts its final position, it is less likely to have an impact on another vehicle compared to higher v/c ratios.
Comparison to current practice

In this section, the ERV intra-link path generated by our proposed system is compared to a currently adopted practice where downstream vehicles go to the nearest edge upon detecting an approaching ERV. In this case, downstream non-ERVs do not act cooperatively and each non-ERV seeks an empty cell on its closest edge after its corresponding minimum stopping distance. The ERV’s optimal intra-link path (which is the same for all tested levels of market penetration) is compared to the one that can be completed by the ERV under the current practice. The results, in Table 4-7, show the considerable ERV travel time reductions in seconds increasing as the v/c ratio increases along an 1827-ft link. Under the current practice, the ERV entering the link segment on the rightmost lane is forced to make a left maneuver to avoid the non-ERVs stopped at the right edge of the road, hence decreasing its speed. In addition, the ERV continues its movement that is inevitably adjacent to stopped non-ERVs on the edges, inhibiting it from increasing its speed. Besides, as non-ERVs are not receiving any assistance and only trying to reach the nearest edge as soon as possible, passing and weaving among vehicles is expected. A risky interaction between two non-ERVs is considered when the two non-ERV are passing each other as they are heading to the same edge. For instance, four, two and eight risky interactions are noted for the scenarios with v/c ratios of 0.75, 0.85 and 0.95 respectively. On the contrary, the proposed approach avoids all types of vehicular interaction, which is by itself an added value. When traditional warning systems exist (sirens and lights), downstream non-ERVs receive limited time to react and interactions between the ERV and non-ERVs can occur, negatively impacting the ERV movement and speeds. For comparison purposes, it is assumed that the presence of an advanced emergency warning system (under the current practice) that notifies the downstream non-ERVs of an approaching ERV early in time so that the non-ERVs are stopped at the edges at the time the ERV arrives, limiting the interaction between the ERV and moving downstream non-ERVs.

Figure 4-7: Percentage of connected non-ERVs that adjusted their instructed final longitudinal position for different v/c ratios and market penetration levels

<table>
<thead>
<tr>
<th>Market penetration level</th>
<th>v/c=0.75</th>
<th>v/c=0.85</th>
<th>v/c=0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of connected non-ERVs that adjusted their instructed final longitudinal position</td>
<td>0.1</td>
<td>0.2</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Table 4- 7: ERV travel time reduction due to our proposed system

<table>
<thead>
<tr>
<th>v/c</th>
<th>ERV travel time improvement (in seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.75</td>
<td>5.04</td>
</tr>
<tr>
<td>0.85</td>
<td>8.33</td>
</tr>
<tr>
<td>0.95</td>
<td>20.47</td>
</tr>
</tbody>
</table>

4.7. Conclusion

In this paper, an approach that optimizes the ERV movement on a link was proposed, extending Hannoun et al.’s (2018) previous work [10]. The previous approach, defining the optimal intra-link path that can be travelled at maximum speed and with maximum adjacent space from non-ERVs after instructing the downstream traffic to stop at specific positions (Hannoun et al., 2018), was impractical on a large link due to large ILP problem size and computation time. Such challenges are overcome in this work by adopting a sequential approach to optimize the ERV movement on larger link segments. The sequential optimization technique consists of applying the ILP on shorter link segments with fewer non-ERVs consecutively, controlling the time overhead of each ILP optimization. In addition, partial market penetration is considered. The system estimates the presence of unconnected vehicles and uses that information in the optimization to reserve space for those present on the link. The estimation technique does not assume the presence of sensors and only uses the connected non-ERVs’ positions. When generating the optimal final position of each non-ERV using the ILP, it is assumed that each estimated, unconnected non-ERV follows the behavior of its respective leader. The space that is virtually reserved for the estimated, unconnected non-ERVs is available for the actual, unconnected non-ERVs’ use.

A set of rules was developed to dynamically identify the minimum feasible stopping range cut-off value needed to avoid infeasibility. Recommendations on how to apply the sequential approach to minimize computation time without compromising the ERV travel time were developed. Smaller IR sizes are preferable at high v/c ratios as long as the reduction in computation time is valuable. In addition, an adequate non-ERV grouping (in IRs) that excludes large gaps between vehicles from IRs can significantly reduce the computation time without sacrificing solution quality.

In addition, the proposed approach was tested at different market penetrations ranging from 70% to 100%. The ERV intra-link path and the connected non-ERV final positions are determined using the ILP executed with the estimated, unconnected non-ERVs. The ERV intra-link path and the connected non-ERV final positions are the information communicated to the ERV and the connected non-ERVs respectively. However, the presence of the actual, unconnected non-ERVs may interfere with the ERV and non-ERV movement. An evaluation technique that consists of re-running the ILP with the actual, unconnected non-ERVs’ positions and with constraints imposing the ERV intra-link path and relaxed final positions of non-
ERVs is performed. Results showed that the generated ERV intra-link path is not affected as long as the connected non-ERVs can slightly adjust their instructed final positions when it is obvious that another non-ERV is obstructing the ERV path.

Our proposed approach was compared to a current practice of going to the nearest edge before the ERV arrival. The optimal ERV intra-link movement generated using the ILPs decreases ERV travel time when compared to the current practice. In the absence of vehicular communications, non-ERVs seeking to reach the nearest edge are likely to undertake risky interactions with other non-ERVs. The proposed approach limits these interactions by prohibiting weaving and passing among non-ERVs.

Future works involve the extension of this approach to make it applicable on a network-wide basis. In addition, multiple emergencies in the network and/or multiple ERVs on the same link at the same point in time will be considered.

Appendix B: ‘tcc’ parameter

The variable $x'_j$ represents the longitudinal index of non-ERV $j$ at the time of data collection in the IR. One of the preprocessing steps consists of determining the minimum stopping distance ($MSD_j$) of each non-ERV $j$ to locate the non-ERV’s minimum final longitudinal index ($x''_j$) with respect to the start of its corresponding AR. In this appendix, the following equation is explained further.

$$MSD_j = ceil \left( \left( tcc \times \sigma_j + t^r \sigma_j + 0.5 \frac{\sigma_j^2}{\delta_j} \right) / L \right)$$

The $MSD_j$ refers to the minimum stopping distance of each vehicle $j$. It is the minimum distance that should be provided between the initial longitudinal position of non-ERV $j$ (at the time of data collection) and its final stopped position. The $MSD_j$ includes (1) the distance travelled during the reaction time ($t^r \sigma_j$) (i.e., distance travelled between the time of instruction receipt and time of braking initiation) and (2) the distance needed to brake from a speed of ($\sigma_j$) to a stop using a comfortable deceleration rate of ($\delta_j$). In addition, $MSD_j$ includes the distance travelled during the communication time and ILP computation time that elapse before the non-ERV $j$ receives the instruction about its final position. During the time of communication and computation called $tcc$, the non-ERV is travelling at its collected speed. The $tcc$ includes the time to communicate the data from the vehicles ($t_{out}$), the computation time needed to generate the optimal solution using the ILP ($CT$) and the time to communicate the instruction to the vehicles ($t_{in}$). Since this is a preprocessing step, the ILP is not yet executed so the $CT$ is not yet known. This is why a lookup table is developed to estimate the $CT$ based on the IR size and the number of non-ERVs.

The $tcc$ is the estimated time between the data collection and the expected time of instruction receipt. The data of $|J|$ non-ERVs is collected at time $t_1$. To limit the ILP problem size, the non-ERVs are
grouped in different IRs. After the execution of the ILP for $IR^1$, the $CT_{\text{actual}}$ of $IR^1$ are identified. If the duration elapsed after the last data collection, which is the sum of $t_{out}$ and $CT_{\text{actual}}$ of $IR^1$, falls below a threshold (use 10 seconds), no data update is needed. Otherwise, a new data collection is ordered at time $t_2$ that is equal to: $t_2 = t_1 + t_{out} + CT_{\text{actual}}$ of $IR^1$.

Note that the threshold of 10 minutes is used but this value is subject and can depend for example on the communication capabilities.

At $t=t_1$ (Data collection of all vehicles)

At $t=t_1+t_{out}+CT_{\text{actual}}$ of $IR^1$ (check: if data update is not needed $\rightarrow$ Option A, if data update is needed $\rightarrow$ Option B)

At $t=t_1+t_{cc}$ of $IR^1$ (Group 1 received the instruction)

$\rightarrow$ Option A: No data update needed
At $t=t_1+t_{cc}$ of $IR^2$ (Group 2 received the instruction while Group 1 may still be braking to final positions)

$\rightarrow$ Option B:
Data update at time $t_2 = t_1+t_{out}+CT_{\text{actual}}$ of $IR^1$

At $t=t_2+t_{cc}$ of $IR^2$ (Group 2 received the instruction while Group 1 may still be braking to final positions)
$\rightarrow$ Option A or B

At $t = t_{\text{final}}$ (Group 1 and Group 2 at final positions and the ERV is approaching)

To sum up, equation (44) is used to determine the $tcc$ used in the preprocessing step for the ILP of $IR^i$:

$$tcc$ of $IR^i =
\begin{cases}
t_{\text{out}} + \sum_{j=1}^{j=i-1} CT_{\text{actual}}$ of $IR^j + CT_{\text{estimated}}$ of $IR^i + t_{\text{in}}$; if $t_{\text{out}} + \sum_{j=1}^{j=i-1} CT_{\text{actual}}$ of $IR^j \leq 10$ seconds (using last data collection) \\
t_{\text{out}} + CT_{\text{estimated}}$ of $IR^i + t_{\text{in}}$; if $t_{\text{out}} + \sum_{j=1}^{j=i-1} CT_{\text{actual}}$ of $IR^j > 10$ seconds (using new data collection)
\end{cases}
$$

where $i \geq 2$ and $IR^1$ is the 1st IR after the last data collection (44)

After completing the ILP for $IR^i$, its actual $tcc$ can be computed after replacing $CT_{\text{estimated}}$ by $CT_{\text{actual}}$ in the equation above. If the estimated $tcc$ is less than the actual $tcc$, the driver will simply arrive earlier to its final position. Otherwise, the $MSD_j$ would have been underestimated because the estimated distance travelled during the computation time was in reality larger. In that case, the non-ERV $j$ may not be able to comfortably reach its final position. To avoid re-solving the optimization, it is assumed that these vehicles are requested to brake at a higher deceleration rate. If the rate is not comfortable, re-solving the optimization while accounting for the lost time of the unused optimization is inevitable. Consequently, overestimating the estimated computation time ($CT_{\text{estimated}}$) is preferred. The estimates needed for are $t_{\text{out}}$ and $t_{\text{in}}$ are obtained using a Python code for round trip time. The $CT_{\text{estimated}}$ is identified using look-up table (Table B-1) based on the size of the IR and the number of non-ERVs.

Note that the size of AR depends on the estimated computation time itself. A good approximation would be the size of the IR or the size of the AR without including the distance travelled during computation time which is the same for all vehicles.
Table B-1: Estimated CT based on the IR longitudinal size and non-ERVs number

<table>
<thead>
<tr>
<th>IR longitudinal size</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
<th>40</th>
<th>45</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.4415</td>
<td>0.5654</td>
<td>0.7134</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.4103</td>
<td>0.6099</td>
<td>0.9009</td>
<td>1.176</td>
<td>1.6693</td>
<td>1.7113</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.4351</td>
<td>0.6267</td>
<td>0.859</td>
<td>1.2646</td>
<td>1.8211</td>
<td>2.3015</td>
<td>3.803</td>
<td>5.2327</td>
<td>6.3925</td>
</tr>
<tr>
<td>20</td>
<td>0.4844</td>
<td>0.7247</td>
<td>1.0738</td>
<td>1.5579</td>
<td>2.2881</td>
<td>3.4586</td>
<td>4.2501</td>
<td>5.6517</td>
<td>7.7978</td>
</tr>
</tbody>
</table>

Appendix C: Imposed Instructions

Due to the possible overlap with the previous AR, the ERV instructions delineating the ERV path up to the most downstream non-ERV in the previous AR are imposed in the next optimization. This way, the ERV does not receive different instructions at increments that are included in more than one AR. Only instructions up to the most downstream non-ERV are disseminated. Examples are given below:

Case 1:
In Case 1, the first ERV instruction shown in AR 1 delineates the ERV movement up to the longitudinal index of non-ERVs v21 and v22. The second ERV instruction shown in AR 1 delineates the ERV movement in an area where no non-ERV from AR 1 are assigned. Only new vehicles from AR 2 are assigned to this area. An overlap between AR 1 and AR 2 exist, resulting in the generation of ERV instruction at the first two increments twice (one in the optimization of AR 1 and second in the optimization of AR 2). Only the first shown ERV instruction from AR 1 is fixed as “Go straight” in the next optimization and disseminated to the ERV right after the generation of the results of the first optimization.

AR 1

<table>
<thead>
<tr>
<th>ERV</th>
<th>ERV</th>
<th>ERV</th>
<th>ERV</th>
<th>ERV</th>
<th>ERV</th>
<th>ERV</th>
<th>ERV</th>
</tr>
</thead>
<tbody>
<tr>
<td>v14</td>
<td>v15</td>
<td>v17</td>
<td>v18</td>
<td>v20</td>
<td>v21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>v16</td>
<td>v19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

straight | straight | straight | straight | straight |
imposed | Not imposed | imposed | imposed | imposed |

AR 2

<table>
<thead>
<tr>
<th>ERV</th>
<th>ERV</th>
<th>ERV</th>
<th>ERV</th>
<th>ERV</th>
<th>ERV</th>
<th>ERV</th>
<th>ERV</th>
<th>ERV</th>
<th>ERV</th>
<th>ERV</th>
<th>ERV</th>
<th>ERV</th>
<th>ERV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>v2</td>
<td>v5</td>
<td>v6</td>
<td>v7</td>
<td>v8</td>
<td>v10</td>
<td>v11</td>
<td>v13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v1</td>
<td>v3</td>
<td>v4</td>
<td>v9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

85
Case 2:
In Case 2, the first ERV instruction shown in AR 1 delineates the ERV movement up to the longitudinal index of non-ERVs v25, v26 and v27. The second ERV instruction shown in AR 1 delineates the ERV movement in an area where one non-ERV (v28) from AR 1 is assigned. Non-ERV v28 is the most downstream non-ERV of AR 1. An overlap between AR 1 and AR 2 exist. In that case, both ERV instructions shown in AR 1 are fixed as “Go straight” in the next optimization of AR 2 and disseminated to the ERV right after the generation of the results of the first optimization. The second ERV instruction shown in AR 1 is fixed even though new non-ERVs from AR 2 may be added along the same increment inclosing v28. If this instruction at this increment can change in AR 2, the ERV may be required to make a left lane change instead of going straight at the second increment of AR 2 and interfere with v28.

**AR 1**

<table>
<thead>
<tr>
<th></th>
<th>v14</th>
<th>v18</th>
<th>v22</th>
<th>v27</th>
<th>v28</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERV</td>
<td>ERV</td>
<td>ERV</td>
<td>ERV</td>
<td>ERV</td>
<td>ERV</td>
</tr>
<tr>
<td>v15</td>
<td>v17</td>
<td>v19</td>
<td>v21</td>
<td>v23</td>
<td>v26</td>
</tr>
<tr>
<td>v16</td>
<td>V20</td>
<td>V24</td>
<td>v25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**AR 2**

<table>
<thead>
<tr>
<th>v1</th>
<th>v3</th>
<th>v6</th>
<th>v7</th>
<th>v8</th>
<th>v9</th>
<th>v11</th>
<th>v12</th>
</tr>
</thead>
<tbody>
<tr>
<td>v2</td>
<td>v4</td>
<td>v5</td>
<td>v10</td>
<td>v13</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Reference


Chapter 5. MESO- AND MICRO-SCOPIC ROUTING OF AN EMERGENCY RESPONSE VEHICLE WITH CONNECTED VEHICLE TECHNOLOGIES

This paper introduces a semi-automated assistance system to improve emergency response vehicles’ (ERV) operations using connected vehicle technologies in an urban transportation network. Much research attention has been given to ERV operations (1) at the network level by providing routing guidance to the ERV and developing efficient fleet dispatching techniques and (2) at the intersection level by granting priority with enough green indication to the ERV’s approach at signalized intersections. The proposed system provides network-wide emergency assistance as it uses routing and intersection preemption techniques. Yet, it also introduces and implements a less-explored aspect of assistance to ERVs which is at the link level. In other words, this system also focuses on facilitating the ERV movement along the transportation links. Confusion experienced by traffic downstream of an approaching ERV is a major issue. Upon hearing the sirens and/or noticing the emergency lights, non-ERVs intuitively yield and try to free space for the ERV. Yet, their combined actions do not always guarantee a safe and efficient passage. For instance, these vehicles may be taking risky maneuvers hence jeopardizing their and others’ safety, or they may inadvertently slow down the passage of the ERV. This paper introduces a system that first determines the optimal ERV route from origin to destination in the network, and second generates assistance messages to the ERV as well as non-ERVs, through the execution of sequential integer linear programs (ILPs) along transportation links. As the ERV is moving towards the destination, new downstream vehicles have to be notified and provided with an instruction message requesting them to stop at a position they can reach comfortably. The ERV receives maneuvering recommendations progressively. The ILP, initially introduced in [1], is extended in this paper to adapt it to the presence of intersections. This proposed system also accounts for partial market penetration of connected vehicle technologies, making it a potential application for short-term deployment. The system is evaluated, using NetLogo, an agent-based modeling tool, in an urban transportation network with different combinations of congestion and market penetration levels. When compared to the case with no system (0% market penetration), results show reductions in terms of ERV travel time when the market penetration level is at a minimum of 40% (average of 9.09% reduction at 100% market penetration) as well as a notable decrease in vehicular interactions (average of 35.46% and 81.38% reduction in ERV/non-ERV and non-ERV/non-ERV reduction respectively at 100% market penetration). It is important to highlight that this system does not result in significant increase in the average time during which the non-ERVs are affected by the ERV’s passage (average of 7.30% increase compared to the case with no system).
5.1. Introduction

Emergency response vehicles’ (ERVs) movements should be facilitated over the complete ERV route from origin to destination. Emergency preemption facilitates the ERV passage at intersections only and emergency warning systems such as lights and sirens are not fully effective [2]. So, in this paper, the ERV passage along transportation links is targeted as it has received little research attention in the past. Hannoun et al. [1] introduce an integer linear program (ILP) that determines, using connected vehicle technologies, the ERV movement along a link segment while generating a specific instruction message to the group of non-ERVs expected to stop on this link segment. The objective of this paper is to deliver a complete system applicable on a network-wide basis. The overarching goal is to develop a comprehensive connected vehicle application that first identifies the optimal ERV route from origin to destination and second facilitates the ERV intra-link movement on this optimal route by instructing downstream traffic to move to specific locations prior to the ERV’s arrival. The proposed system identifies the ERV optimal route using the hyperstar routing algorithm. Next, a criticality analysis module is developed to determine when and which group of downstream non-ERVs to consider in the next ILP optimization. The ILP, introduced in [1], is extended so that it can be applied on larger links by adopting a sequential approach in which the adjusted ILP is executed on consecutive link segments. In addition, it is adjusted to accommodate partial market penetration levels (i.e., presence of unconnected downstream non-ERVs unable to share data or receive instructions). This paper also extends the previous work by accommodating the presence of intersections. The system is evaluated on a sample transportation grid network under different initial conditions using the NetLogo agent-based modeling environment. Results in terms of ERV travel times and vehicle interactions are investigated in detail and compared to a currently adopted practice for different market penetration levels. The following section (Section 5.2.) includes a literature review about previous studies that targeted emergency response operations. Next, the system’s general approach is proposed (Section 5.3.) followed by a detailed description of each of its module (5.4. through 5.6). The experimental analysis plan and results are discussed in Section 5.7. and Section 5.8. respectively. Finally, Section 5.9. presents the key finding and future works.

5.2. Literature Review

Emergency response systems are indispensable and the reliability of their services can directly affect people’s lives. A recent study in Scotland highlighted that the majority of trauma fatalities occur before hospital care [3], making the efficiency of emergency response systems critical. The need for further enhancements of emergency response systems’ operations will never cease to grow. The emerging capabilities of ITS (i.e. Intelligent Transportation Systems) technologies brought considerable benefits to urban transportation networks and to emergency response systems in particular. For example, automatic
crash notification systems mainly relying on the transmission of sensor data have resulted in reduced ERV response time and mortality [4-6]. In case of an incident, these systems automatically contact the closest emergency call center and send information relevant to dispatching and staging decision-making. As part of the US-DOT’s Dynamic Mobility Applications (DMA), R.E.S.C.U.M.E (i.e. Response, Emergency Staging and Communications, Uniform Management and Evacuation) is a bundle that also includes an application for automatic crash notification (AACN) as well as other applications for assistance at incident scenes (INC-ZONE), during evacuation situations (EVAC) and for dispatching/staging guidance through improved situational awareness (RESP-STG) [7].

From a transportation engineering perspective, researchers have primarily focused on improving emergency response systems at the intersection level and at the network level. Intersections are major conflict points where most of the ERV-related crashes occur [8]. Emergency preemption is a type of preferential treatment granted to ERV’s at signalized intersections for faster response and enhanced safety. It consists of prioritizing the ERVs’ movement by altering traffic signal phasing (i.e., providing enough green indication to the ERV and red indication to conflicting approaches) [9]. The emergency preemption concept is already widely deployed at signalized intersections in the United States [10]. With the advancements in ITS, the technologies used for emergency vehicle detection are being upgraded. Furthermore, the shift from an intersection-by-intersection basis to a more dynamic route-wide emergency preemption is explored [9, 11, 12].

On the other hand, routing and managing a fleet of emergency vehicles in a network requires highly adaptable systems as no information about upcoming emergency requests (demand) can be obtained (with certainty) in advance to pre-distribute resources [13]. To improve the emergency response system’s coverage of demands and hence response times, systems for dynamic dispatching and relocation of idle emergency vehicles have been developed [14-16]. As for the routing of a single ERV from an origin to a destination, general (i.e., not ERV specific) dynamic shortest-path algorithms are usually employed [16]. The use of a routing algorithm is necessary to the system proposed in this paper, so that intra-link movement assistance is generated and disseminated to the ERV and non-ERVs along the selected ERV route from origin to destination. The time-dependent vehicle routing algorithm called Hyperstar developed by Bell et al. [17] is identified as a good fit for the proposed system. First, it considers the dynamic aspect of urban transportation networks by assuming that link performance measures are not static. For instance, they vary with the time of the day and are subject to uncertainty due to irregular incidents. Second, the hyperstar routing algorithm generates, using a goal-directed search, a set of alternative routes, called a hyperpath, leaving the final route selection within the ERV driver’s discretion (further discussed further in Section 5.4.). In this paper, a single ERV is considered. Subsequently, strategies for efficient deployment of an ERV
fleet in the urban transportation network are not included in this proposed system. This will be addressed in future works along with routing adjustment recommendations when multiple emergencies are considered.

Studies that are the most relevant to the proposed system are the ones addressing the movement of ERVs along transportation links. Currently, the techniques used by ERVs to facilitate their passage along streets are strobe lights and sirens [9]. Yet, these traditional warning systems do not fully eliminate confusion and risks of vehicle collisions [2]. For instance, a non-ERV may still be unsure about how to efficiently react upon detecting an ERV. It may end up obstructing the ERV’s movement if it did not free the ERV’s way on time or may succeed in doing so while jeopardizing its and others’ safety due to risky maneuvers. A stream of research has been conducted to develop more advanced emergency warning systems [18-20]. These systems aim to provide surrounding non-ERVs with early warning messages along with relevant information to ensure that downstream traffic is well-aware of the incoming ERV and is granted enough time and details to react adequately. Yet, the decision about how to react is left to the discretion of the drivers which means that timely and proper actions are not guaranteed. Systems providing downstream traffic with early warning messages that also enclose advice about how to maneuver exist but are limited. Buchenscheit et al. [21] presented a conceptual prototype of an advanced emergency warning system that gives ERV route information and maneuver recommendations to downstream vehicles by relying on inter-vehicle communication. They evaluated the need for this additional safety measure using an expert survey which showed high acceptance. Inspired by Buchenscheit et al. [21], Weinert et al. [22] develop a system that first preempts the passage of an ERV at a signalized intersection using V2I (vehicle-to-infrastructure) communications and second creates a rescue lane for the ERV’s passage using V2V (vehicle-to-vehicle) communications. For the second application, a static rule-based approach is adopted to generate the warning messages and determine whether the non-ERVs should change lanes. This approach assumes a predefined ERV intra-link path and assumes that the non-ERVs that are travelling on the ERV’s lane are the only ones of interest (i.e., they need to move). Nevertheless, a more dynamic intra-link ERV path that is generated based on the downstream traffic may lead to better travel time. In other words, an ERV steering away from a platoon of vehicles may reach better and more reliable ERV travel time along with reduced non-ERV interactions than an ERV maintaining a pre-determined movement and expecting the downstream non-ERVs to cooperate and create a rescue lane. Another system developed by Yoo et al. [23] consists of requesting downstream non-ERVs to shift away from the lowest density lane to reserve it for ERV use. Yet, this system needs to be extended to account for desired lanes at upstream and downstream intersections, as the lowest density lane may not always be the most optimal and adequate one for ERV use. This paper presents a system that generates the optimal ERV intra-link movement based on the downstream traffic by minimizing vehicular interactions, so the ERV intra-link movement is not pre-defined. The system generates optimal maneuvering recommendations to the ERV as well as non-ERVs instead of following a
set static rules that may not be feasible (or safe). Also, the system takes into consideration the desired lane occupancies prior to each signalized intersection depending on the movement to be executed at that intersection (based on the selected route to destination) to ensure a safer and more comfortable trip.

5.3. Proposed system

This paper introduces a centralized system that collects and processes data from connected vehicles (ERV and non-ERVs) then generates and disseminates messages back to them. The proposed system consists of three modules, as shown in Figure 5-1. First, the ERV route generation module determines the optimal route using the time-dependent hyperstar routing algorithm [17]. The second module is the Criticality Analysis module (CAM) that screens the downstream non-ERVs as the ERV is travelling towards its destination to identify the group of *critical* vehicles that are expected to conflict with the ERV’s path. These are the vehicles that should be notified about the approaching ERV and informed about how to effectively react at the earliest time possible. This module filters the non-ERVs to avoid generating and sending unnecessary instruction messages to vehicles that are expected to diverge from the ERV’s route at intersections before the ERV’s arrival. It is only when a critical vehicle is identified by the CAM that the third module, called the Sequential Integer Linear Program optimization module (SIOM), is activated. The latter adopts the sequential approach (discussed in Chapter 4) that consists of executing the ILP, introduced in [1] (Chapter 3). This module preprocesses the vehicle data collected from the connected non-ERVs identified by the CAM and executes an ILP. By maximizing the ERV speeds and the free space around its movement, the mathematical program (1) optimizes the ERV’s intra-link movement along the downstream road segment where the critical non-ERVs can stop, (2) determines the best ERV maneuvering actions to take along that segment and (3) identifies where each non-ERV should stop before the arrival of the ERV. The SIOM adopts a sequential approach because the ILP is optimizing the intra-link movement of the ERV along short link segments on the selected ERV’s route sequentially as the ERV is travelling and as new vehicles become critical. One of the preprocessing steps in this module considers the output of previous ILP optimizations to ensure continuity of the ERV movement and consistency of the instruction messages being disseminated to the ERV and non-ERVs. This module also considers the presence of signalized intersections and infers based on the upcoming movement at the intersection, the desired lane to be occupied by the ERV on the upstream and downstream link of the intersection to ensure a smooth ERV passage. In addition, as the transition to a fully connected vehicle environment is expected to take decades [24], it is crucial to account for partial market penetration conditions in this proposed system. This third module applies an estimation technique as a pre-processing step prior to solving the ILP. It predicts the presence of unconnected non-ERVs between pairs of connected non-ERVs based on a distance criterion. These
additional non-ERVs considered in the ILP virtually reserve extra spots for the actually present unconnected non-ERVs.

Data flows between the proposed centralized system, the ERV and non-ERVs are shown in Figure 5-2. The one-time data flow occurs once between entities. The periodic data flow occurs periodically (for example 0.5 second). The “as-needed” data flow is a data exchange that occurs whenever new information or an update is available (i.e., when the ERV intra-link movement or the non-ERV instruction are generated). In this paper, it is assumed that all vehicle-to-infrastructure communications (i.e., ERV/non-ERV to proposed centralized system) are assured. Assessing the implications of imperfect communications is part of future works, as a step forward towards deployment. Each of the modules is discussed thoroughly next.

![Figure 5-1: Proposed system composed of three modules](image-url)
5.4. ERV route generation module

The expected arrival time’s reliability is a major concern for users of navigation systems. Generating a single path from origin to destination by excluding all unreliable links can result in a path that may still be unreliable possibility because it is not easy to drive (full of turns). This is why it is important to consider the generation of a set of good alternative paths especially as congestion materializes with time and after a trip starts. The time-dependent hyperstar algorithm introduced by Bell, Trozzi [17] identifies a set of alternative optimal paths, called hyperpaths, from origin to destination by minimizing the expected arrival time to the destination and to any intermediate node. The hyperstar algorithm allows driver’s preference. This is very important to ERVs’ drivers, who are expected to be familiar with the area they are serving, as they may select their preferred route based on their preference and/or previous experience. Due to the large number of links in transportation road networks, efficient path finding algorithms are highly recommended and this algorithm incorporates a goal-directed search to accelerate the computation. This algorithm is regarded as a hyperpath version of the time-dependent A* algorithm, which is a speed up of Dijkstra’s classic shortest path algorithm. The time-dependent hyperstar algorithm uses two dynamic link performance measures: the undelayed travel time (varying with different times of day), and the maximum delay experienced due to unexpected events. This algorithm captures the dynamic aspect of road networks by assuming that link performance measures used for the path finding algorithm are not static. They vary with the time of the day and are subject to uncertainty due to irregular incidents. Depending on the available data sources, the undelayed link travel time can be obtained from historic data, sensor data and/or data from connected vehicles. The maximum delay can be assigned to a link due to unexpected events such as vehicle crashes.
This vehicle routing algorithm is considered a good fit to the proposed system. Yet, it can be replaced by any other routing algorithm. The major contributions of this paper are the CAM and SIOM modules (discussed in Sections 5.5 and 5.6, respectively) which depend on the selected ERV route to destination but not on the technique used to generate that route.

5.5 Criticality Analysis Module

The CAM identifies the next set of downstream non-ERVs which should receive an instruction message. In other words, as the ERV is moving, the CAM defines the next group of downstream non-ERVs, whose real-time information should be processed by the Sequential ILP Optimization Module. The latter generates a specific instruction message for each vehicle enclosing a final position to be reached before the ERV passage. Subsequently, the CAM determines a dynamic detection range (discussed below) measured from the ERV’s current position. This range is defined in a way to allow a once detected vehicle enough time to react and execute the instructions before the ERV’s arrival. Otherwise, the proposed system would be inefficient as the ERV would have to interact with downstream vehicles which are still aiming to reach their assigned positions. After determining the adequate detection range, the CAM identifies the non-ERVs of interest (i.e., located in this range) and sends their real-time information to SIOM. The CAM is executed regularly as the ERV is moving towards its destination. The smaller the time step (i.e., CAM’s execution frequency) supported by the available wireless communication network, the better.

5.5.1 CAM Methodology

The methodology that the CAM uses is based on a heuristic approach that consists of identifying the following groups of non-ERVs:

- **Non-ERVs of interest**

  These are the connected non-ERVs that will potentially interact with the ERV if no instruction message is sent to them. They are downstream of the ERV and on its remaining route. They are also the non-ERVs that have not yet received an instruction message from the system (i.e., unassisted non-ERVs). Note that non-ERVs travelling on the opposite direction of the ERV route are not considered as contraflow operation is not in the scope of this dissertation. Regarding the non-ERVs inside intersections, the CAM identifies the ones heading to a link in the ERV’s route based on data about their compass direction. Based on data about emergency preemption detection ranges [25], the emergency preemption has a higher detection range than the system’s detection range which is discussed next. So, the CAM does not consider the vehicles that have not entered the ERV’s route yet as it assumes that emergency preemption at signalized intersections prohibits these non-ERVs from entering the intersection and hence joining the ERV’s route. New vehicles join the ERV route at intersections where preemption is not yet activated. But these non-ERVs are expected to be considered later, when the ERV gets closer to them.
Critical non-ERVs

From the set of non-ERVs identified previously, the ones within a distance of \((\Delta d_{det} + \epsilon)\) from the current ERV’s position are identified. \(\Delta d_{det}\) is the minimum detection range that includes the critical non-ERVs (i.e., whose real-time information should be processed immediately so that the non-ERV receives the instruction and stops before the arrival of the ERV). \(\epsilon\) is a relatively short distance that is added to \(\Delta d_{det}\) to include the non-ERVs that may become critical during the CAM execution’s time step. \(\Delta d_{det}\) and \(\epsilon\) are computed using Equations (1) and (2) respectively that are developed based on time-space diagrams (see Appendix D for more details).

\[
\Delta d_{det} = FS[(v_e - v_v) * (\Delta t_{cc} + t_r + \frac{v_v}{g}) + \frac{v_v^2}{2g}]
\]

\[
\epsilon = \Delta t (v_e - v_v)
\]

Where: 

- \((v_e)\) is the average non-ERV speed (ft/sec);
- \((v_v)\) is the current ERV speed (ft/sec);
- \((t_r)\) is the non-ERV’s reaction time (sec);
- \((\Delta t_{cc})\) is the communication and computation time interval (sec), which is a function of the number of non-ERVs and the largest longitudinal distance between a pair non-ERVs;
- \((g)\) is the non-ERV’s deceleration (ft/sec²);
- \((\Delta t)\) is the time-step of the CAM; and
- \((FS)\) is a factor of safety to account for additional time needed for non-ERVs to change lanes.

Additional non-ERVs

In this step, the group of critical non-ERVs is expanded to add the ones that may become critical and within a distance of \((\Delta d_{det})\) during the time of communication and computation of the next ILP optimization. This is why these non-ERVs should also be considered in the next ILP optimization. These non-ERVs are beyond the critical non-ERVs and within a distance equal to \((\Delta d_{det} + \epsilon + \Delta d_{add1})\) downstream of the ERV (see Appendix D for more details), where \(\Delta d_{add1}\) is determined as follows using Equation (3):

\[
\Delta d_{add1} = \Delta t_{cc}(v_e - v_v)
\]

If more non-ERVs can be added without affecting the maximum allowable communication and computation time for the next ILP optimization, then new non-ERVs are added only if their estimated waiting time at a final position does not exceed a maximum tolerated duration \((\Delta t_{wait})\). These non-ERVs are the ones that have not been included yet and within a distance of \((\Delta d_{det} + \epsilon + \Delta d_{add2})\) from the ERV’s current position (see Appendix D for more details), where \(\Delta d_{add2}\) is determined using Equation (4):

\[
\Delta d_{add2} = v_e\Delta t_{wait}
\]
The complete set of non-ERVs to undergo the next sequential ILP optimization and to receive instruction messages consists of the critical non-ERVs and additional non-ERVs groups.

5.5.2) Filtering of vehicles

The CAM screens the vehicles that are on the ERV’s remaining route to detect the critical vehicles that need to be notified, as discussed. As for the non-ERVs inside the intersection, it only scans the ones heading to a link in the ERV’s route.

Due to the presence of intersections, some vehicles may diverge from the ERV’s route after the data collection and before the receipt of any instruction. In addition, it is not recommended to ask a non-ERV to brake and reach a final position if its minimum stopping distance only allows it to stop beyond an intersection at which it was planning to exit the ERV’s route. This non-ERV is not required to stop and to follow any instruction generated by the system. Subsequently, filtering the groups of non-ERVs determined in Section 5.5.1) is executed. In other words, to avoid having vehicles following an instruction and stopping unnecessarily, this filtering technique determines whether a non-ERV should be considered in the next ILP optimization, before adding this new non-ERV to the next optimization due to its presence within $(\Delta d_{det} + \epsilon + \Delta d_{add1})$ or $(\Delta d_{det} + \epsilon + \Delta d_{add2})$. A parameter $\varphi_j$ is defined for each non-ERV $j$ and takes the value of 1 if $j$ has to be excluded and 0 otherwise.

If the non-ERV is initially positioned on a link in the ERV’s route and upstream of an intersection and its minimum final position ($MFP_j$) defined based on its minimum stopping distance ($MSD_j$) is beyond the intersection’s stop line and if the non-ERV is expected to remain on the ERV’s route, then the non-ERV should receive an instruction and $\varphi_j = 0$. Otherwise, the non-ERV can continue its movement without stopping for the ERV and $\varphi_j = 1$. For example, a non-ERV is excluded if it cannot stop before the intersection at which it intends to make a left-turn while the ERV is following a through movement.

Requesting the connected non-ERVs to share their future movements at intersections is challenging, this is why the intended movement is deduced based on the non-ERV’s lateral position as non-ERVs that are about to enter the intersection are not expected to change lanes anymore.

Consider a 4-way intersection where each approach is composed of 3 lanes with one traversable right shoulder, a right lane for through movements or right turns, a middle lane for through movements and a left lane exclusively for left turns. If the ERV is performing a through movement at that intersection, the detected non-ERVs that are before the stop line, on the leftmost lane and only capable of stopping after the intersection (due to the need of a minimum stopping distance) are excluded from the next ILP optimization as these vehicles will diverge from the ERV’s route before the latter’s arrival. However, the non-ERVs that are located on the middle lane or rightmost lane are expected to remain on the ERV’s route, and as a result are included in the next ILP optimization. Note that a non-ERV on the rightmost lane may not
necessarily perform a straight movement and may diverge from the ERV’s route by making a right-turn. Yet, assuming the worst-case scenario is preferred and this non-ERV will not be excluded. In case of diversion from the ERV’s route, a new message will be generated to drop its instruction after it exits the ERV’s route. For instance, it is always possible to fail in excluding a non-ERV that is in fact exiting the ERV’s route. Even if the \( MFP_j \) is prior to the intersection, non-ERV \( j \) may be assigned to a location in its FSR that is in fact after the stop line of an intersection at which it exited the ERV’s route. Subsequently, a post-processing step is added before the dissemination of the non-ERV’s instruction that determines whether each non-ERV should receive the instruction or not based on its final assigned position with respect to the intersection and its expected movement.

To summarize, for a 3-lane link with the leftmost lane for left-turn, middle lane for through movement and rightmost lane for through movement or right turn, \( \varphi_j \) takes the value of 1 if the non-ERV \( j \) is upstream of an intersection at the time of data collection and cannot stop before the intersection and if:

- It is on the leftmost lane and the ERV is making a through or right movement inside the intersection
- It is on the middle lane and the ERV is making a left or right movement at the intersection
- It is on the rightmost lane and the ERV is making a left movement at the intersection.

To limit interaction between non-ERVs, it is important to make sure that a non-ERV that is excluded from the ILP due to its inability to stop before the intersection at which it is exiting the ERV’s route, is not indirectly forced to overtake a downstream non-ERV that is going to receive an instruction because its speed allows it to stop prior the intersection. In this situation, the non-ERVs that are directly downstream (same lane) of a non-ERV with \( \varphi_j = 1 \) are excluded as well.

The filtering technique is crucial as it may also exclude all critical non-ERVs detected within a distance of \( (\Delta d_{det} + \varepsilon) \) downstream of the ERV. In that case, no ILP optimization is needed. In addition, it limits the optimization problem’s size and, as a result, the computation times.

The CAM and SIOM assume that non-ERVs’ speeds are maintained after data collection. So, after identifying the final set of non-ERVs for the next ILP optimization, a warning message is sent to these vehicles requiring them to preferably maintain their speeds and to expect an instruction message shortly. This will primarily limit the risk of assigning non-ERVs to a final position that cannot be reached (position upstream of their actual feasible stopping range). In addition, this first warning message, which is the same for all non-ERVs detected and filtered by the CAM, will lead to enhanced reaction times.

5. 5. 3)  Start of ERV movement

ERVs are assumed to be parked at a station at the time the emergency call is received. The turnout time is the duration elapsed between the time at which the station is notified of an emergency until the time
at which the ERV leaves the facility. According to NFPA 1710 [26], the turnout time for fire emergencies should not exceed 80 seconds, compared to 60 seconds for EMS incidents. Based on the emergency notification time and turnout time, the desired time of ERV departure can be estimated and is considered input to the proposed system.

Since the focus of this research is not on the optimization of the emergency preemption of traffic signal control, it is assumed that emergency preemption at the next intersection is activated on time and prior to the ERV departure’s time to allow the discharge of the stopped queue on the ERV approach so that the ERV does not approach non-ERVs that are still stopped at the intersection.

The proposed system sends instruction messages to downstream traffic. A vehicle that receives a message should follow the instruction and reach the assigned final position before the ERV’s arrival. This is why new vehicles are prohibited entry to the first link of the ERV’s route for a short time ($\Delta t_{XW}$) prior to the ERV’s desired departure time using traffic control upstream of the ERV’s origin. Consequently, at the desired time of ERV departure, the non-ERVs present on the first link are at the minimum detection distance of $(\Delta d_{det} + \varepsilon)$ from the ERV’s initial position. $\Delta t_{ne}$ is computed as follows using Equation 5 for more details):

$$\Delta t_{ne} = \frac{\Delta d_{det} + \varepsilon}{v_o}$$  \hspace{1cm} (5)

Where $(\Delta d_{det} + \varepsilon)$ is the minimum and additional detection distance as computed in Section 5. 5. 1) and $(v_o)$ is the average non-ERV speed (ft/sec).

5. 6. Sequential ILP Optimization Module

The sequential ILP optimization module is an extension to the approach presented in [1] (Chapters 3 and 4) that is applicable to transportation links without considering the presence of intersections at which non-ERVs can diverge from the ERV’s route. This modified approach relies on input data deduced from the ERV route generation module (discussed in Section 5. 4.) and CAM (discussed in Section 5. 5.). In this subsection, the system that was previously introduced is briefly described with its aspects that remained unchanged. In addition, the new extensions that allows it to be applicable in an urban transportation network are discussed.

5. 6. 1) System description

The ILP is developed to identify the best ERV intra-link passage based on downstream traffic by maximizing the ERV speed and the free space adjacent to its movement. The ILP generates assistance messages to the ERV about the next maneuvers to take as well as a unique warning/instruction message for each downstream non-ERV, requesting it to stop at a specific final position along the link. As the ERV is travelling towards its destination, new non-ERVs become critical. For instance, it is not recommended to
apply the ILP a single time and optimize the positions of all non-ERVs that are present on the remaining route of the ERV all at once because this will force them to stop extremely early and unnecessarily in case they were intending to diverge from the ERV’s route before the latter’s arrival. Subsequently, in this paper, the non-ERVs of interest are the ones identified in CAM (discussed in Section 5.5). Data about the position and speed of each of the non-ERVs is needed as input to the ILP to ensure that a non-ERV is assigned to a position it can safely reach. A preprocessing step (discussed in Chapter 4) estimates the presence of unconnected vehicles to accommodate partial market penetrations. Non-ERVs are assumed to be travelling along an initial range (IR) at the time of data collection. The feasible stopping range (FSR) of each non-ERV is determined to identify the optimization range, that we call the assignment range (AR) along the link. For instance, the ILP generates the ERV intra-link movement and non-ERV assignment positions along this AR. Next, as the ERV is travelling, a new ILP execution is required with new downstream non-ERVs. The corresponding AR location along the link of this new ILP depends on the FSR of the new non-ERVs and may overlap or form a gap with the previous ARs, as shown in Figure 5-3. Subsequently, the output of previous optimizations should be considered to infer input data, to ensure continuity of the ERV movement and to avoid any confusion related to assigning more than one non-ERV to the same location.

![Initial ranges (IRs) and corresponding assignment ranges (ARs)](image)

*Figure 5-3: Initial ranges (IRs) and corresponding assignment ranges (ARs)*

The ILP presented in [1] (and Chapters 3 and 4) is applicable to a transportation link. For instance, all detected non-ERVs are on the same transportation link and are expected to remain on it after data collection and until they complete their corresponding instruction. The optimization space (AR) does not include an intersection in which non-ERVs cannot stop and after which non-ERVs may diverge from the ERV’s route. Subsequently, the proposed system previously presented in [1] needs to be adjusted to become applicable to a transportation network with intersections. Detected non-ERVs may be on different transportation links of the ERV route at the time of detection or may be assigned to final locations on different links of the ERV route. Hence, non-ERVs may have different headings in the IR and/or in the AR, as shown in Figure 5-4. Heading is a number in degrees between 0 and 359, measured from the north in a clockwise direction (i.e., 0 is north, 90 is east).
As presented in [1], the system discretizes a transportation link into identical cells. The x-axis represents forward motion while the y-axis represents lateral motion. To ensure consistency of the system’s setup, the X and Y axes rotate with the different headings, as shown in Figure 5- 5.

Non-ERVs cannot be assigned to positions inside the intersections. The ERV has a predefined movement inside the intersection that depends on the next link in its optimal route. Since the intra-link movement inside the intersection is fixed (i.e., does not need to be generated by the ILP) and since non-ERVs are not supposed to stop in the intersection area, there is no need to include the intersection space in the ARs (i.e., optimization spaces) and to unnecessarily increase the problem size and computation time. Figure 5- 6-a through Figure 5- 6-e describe how the ERV route in an example transportation network
(highlighted in Figure 5- 6-a) is pictured by the sequential optimization module: as a set of multiple links with the same or different headings and without intersections (Figure 5- 6-e).

![Figures (a) through (e) show how the links are virtually rotated for X and Y axis’ consistency and the intersection space is excluded.](image)

Since the sequential ILP optimization omits the intersection space, it places the non-ERVs, that are in reality initially positioned inside the intersection at the time of detection, on the stop line prior to that intersection. Yet, the lateral position on the stop line would depend on the movement being performed inside the intersection (Figure 5- 7). For instance, as a preprocessing step prior to the activation of the ILP, each non-ERV receives an ID that increases with the distance to the ERV and with the lateral position, regardless of whether it is positioned on a link or in the intersection. It is due to this ID that the ILP infers the true positions of the vehicles that may look overlapping. The ILP subsequently generates an output that ensures that there is no weaving among non-ERVs and that no more than one non-ERV is assigned to a cell along a link.
Figure 5-7: Non-ERVs inside intersection at detection are considered positioned on the stop line.

Depending on the optimal ERV route generated by the first module described in 5.4, the desired lateral positions upstream and downstream of each intersection in the ERV route are deduced based on the movement to be performed inside the intersection. For example, the ERV should occupy the leftmost lane before entering an intersection in which it will make a left movement. The following section 5.6.2) presents the notation of the variables and parameters used in the preprocessing steps and mathematical program described in Sections 5.6.3) and 5.6.4), respectively.

5.6.2) Notation

Table 5-1 and Table 5-2 show the parameter and variable notation, respectively.

Table 5-1: Parameter notation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>Longitudinal size of a cell (in ft)</td>
</tr>
<tr>
<td>$W$</td>
<td>Lateral size of a cell (in ft)</td>
</tr>
<tr>
<td>$N$</td>
<td>Longitudinal size of an ERV (in cells)</td>
</tr>
<tr>
<td>$LL$</td>
<td>Longitudinal size of the AR (in cells)</td>
</tr>
<tr>
<td>$Y$</td>
<td>Lateral size of the AR, with traversable shoulders (in cells)</td>
</tr>
<tr>
<td>$J$</td>
<td>Number of non-ERVs</td>
</tr>
<tr>
<td>$AR_{start}$</td>
<td>Distance between the start of the AR and its IR (in cells)</td>
</tr>
<tr>
<td>$x^{AR}_{sl}$</td>
<td>Longitudinal index of the next intersection’s stop line with respect to $AR_{start}$</td>
</tr>
<tr>
<td>$t^r$</td>
<td>Reaction time of non-ERVs (in seconds)</td>
</tr>
<tr>
<td>$\sigma_j$</td>
<td>Initial speed of non-ERV $j$ (in fps)</td>
</tr>
<tr>
<td>$\delta_j^{max}$</td>
<td>Maximum comfortable deceleration rate of non-ERVs (in fps$^2$)</td>
</tr>
<tr>
<td>$t_{cc}$</td>
<td>Communication and computation time (in seconds)</td>
</tr>
<tr>
<td>$x_j$</td>
<td>Initial longitudinal index of non-ERV $j$ with respect to the start of the IR</td>
</tr>
<tr>
<td>Notation</td>
<td>Description</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>$y_j'$</td>
<td>Initial lateral index of non-ERV $j$</td>
</tr>
<tr>
<td>$MSD_j$</td>
<td>Minimum stopping distance of non-ERV $j$ (in cells)</td>
</tr>
<tr>
<td>$MFP_j$</td>
<td>Minimum final longitudinal index of non-ERV $j$ with respect to the start of the IR</td>
</tr>
<tr>
<td>$x_j''$</td>
<td>Minimum final longitudinal index of non-ERV $j$ with respect to the start of the AR</td>
</tr>
<tr>
<td>$c$</td>
<td>Longitudinal Feasible Stopping Range cutoff value (in cells)</td>
</tr>
<tr>
<td>$S_{\text{free}}$</td>
<td>Maximum allowable speed of the ERV (in speed stage)</td>
</tr>
<tr>
<td>$np^{x,y}$</td>
<td>Binary parameter that takes the value of 1 if cell $(x,y)$ cannot be occupied by a non-ERV to ensure the no passing rule across ARs and 0 otherwise</td>
</tr>
<tr>
<td>$d_k^{i,y}$</td>
<td>Binary parameter that takes the value of 1 if instruction $k$ was previously sent to the ERV at increment $i$ and lateral index $y$ and 0 otherwise</td>
</tr>
<tr>
<td>$su^i$</td>
<td>Integer parameter that takes the value of $s_{env}^i$ if this variable was generated previously at increment $i$ due to overlap with previous ARs and ($S_{\text{free}} + 1$) otherwise (see Table 5-2 for $s_{env}^i$ description)</td>
</tr>
<tr>
<td>$y_d$</td>
<td>Desired ERV lateral position</td>
</tr>
<tr>
<td>$\alpha^i$</td>
<td>Binary parameter that takes the value of 0 if cells forming a stop line exists at $x \geq (N+1)i$ and $x &lt; (N+1)(i+1)$ and 1 otherwise, for $i = 1, \ldots, LL/(N+1) - 1$</td>
</tr>
<tr>
<td>$\alpha^f$</td>
<td>Binary parameter that takes the value of 0 when an ERV lateral position is desired at the end of the AR due to an upcoming intersection (within 2 increments after the AR) and 1 otherwise</td>
</tr>
<tr>
<td>$ty_j$</td>
<td>Binary parameter equal to 1 if non-ERV $j$ is a connected vehicle and 0 otherwise</td>
</tr>
<tr>
<td>$le_j$</td>
<td>Integer parameter equal to the label of the leader of each non-ERV $j$ in the IR and equal to $j$ if the leader of non-ERV $j$ is not included in the IR</td>
</tr>
</tbody>
</table>

Table 5-2: Variable notation
5.6.3) Preprocessing steps

- **Unconnected non-ERV estimation**

  This system accommodates the presence of unconnected non-ERVs that are unable to share their real-time information and receive instruction messages when an ERV is approaching. Upon receiving the connected non-ERVs’ data, a preprocessing step estimates the presence of unconnected non-ERVs in the IR. The ILP is independent of the estimation technique. In this research, it is inspired by [27] and unconnected non-ERVs are added between pairs of connected non-ERVs using a distance criterion based on the Wiedemann car-following model for the slow-down region. Predicting the position of the unconnected non-ERVs is one step towards the accommodation of partial market penetration in the proposed system. Constraints are added to the ILP to differentiate between a connected and unconnected non-ERV. While the former is expected to follow the received instruction message, the latter’s movement is estimated by assuming a “follow the leader” behavior using a constraint (Equation 49). Two binary parameters are needed for this constraint. The first parameter is \( ty_j \) which refers to the type of the non-ERV by taking the value of 1 for a connected non-ERV and 0 for an unconnected non-ERV. The second parameter is \( le_j \) which refers to the leader of non-ERV \( j \). For connected non-ERVs, \( le_j \) can refer to the leader of non-ERV \( j \) even though the constraint (Equation 49) will not be binding due to \( ty_j \) equal to 1. If the leader of non-ERV \( j \) is not included in the same ILP optimization, \( ty_j \) is set to \( j \).

- **Feasible Stopping Range (FSR) identification of each non-ERV \( j \)**

  This step consists of determining, for each connected non-ERV \( j \) detected by the CAM and unconnected non-ERV added in the previous step, a FSR within which non-ERV \( j \) can comfortably stop. For instance, a non-ERV should never be requested to stop at a position that it cannot reach, i.e., at a distance smaller than its minimum stopping distance (\( MSD \)). Subsequently, the \( MSD_j \) of each non-ERV \( j \) is determined using Equation 6. It includes (1) the distance travelled during the computation time (\( tcc \)) and during the reaction time (\( t^r \)) and (2) the braking distance using a maximum comfortable deceleration of (\( \delta_j^{max} \)). The (\( MSD_j \)) is computed in terms of cells and cannot be smaller than 1 cell. This is to make sure that a non-ERV at zero speed does not receive an instruction to shift lanes without moving forward. After determining the (\( MSD_j \)) of each non-ERV \( j \), the minimum final position (\( MFP_j \)) with respect to the start of the IR is identified using Equation 7.

\[
MSD_j = \max \left( 1, \frac{\left[(tcc+t^r)\sigma_j + \frac{\sigma_j^2}{2\delta_j^{max}}\right]}{L} \right) \quad (6)
\]

\[
MFP_j = x_j' + MSD_j \quad (7)
\]
The ILP optimizes the ERV intra-link movement and non-ERVs final positions over a link segment that is called AR. The AR is a range along the link segment that includes the FSR of all the non-ERVs in the current optimization. The reason why an AR is defined is to limit the problem size because there is no need to execute the optimization over a space which no non-ERV can utilize. The AR starts at least one increment prior the smallest $MFP_j$ and ends at or after the largest $MFP_j + c$ (where $c$ is the FSR size) in a way to make the AR size a multiple of the ERV’s longitudinal size + buffer. After defining the start of the AR ($AR_{start}$) with respect to the start of the IR, the minimum final index ($x_{j''}^{+}$) of each non-ERV in the AR is determined, using Equation 8.

$$x_{j''}^{+} = MFP_j - AR_{start} + 1$$  \hspace{1cm} (8)

The FSR of each non-ERV starts at its $x_j^{'}$ and extends a number of $c$ cells beyond it. This is to limit the problem size and to make the computation time smaller by reducing the space within which each non-ERV can stop. The $c$ parameter is an input to the ILP and is the same for all non-ERVs (this can be relaxed in the future). When no feasible solution is found, re-executing the problem with a higher $c$ may lead to feasibility. In other words, as $c$ increases, new combinations of non-ERVs’ final positions are added, potentially leading to a combination satisfying all constraints yet resulting in higher computation times. The computation time includes the duration spent while increasing $c$ until a feasible solution is found.

Subsequently, two rules of thumb are developed to determine the minimum $c$ value that should be used initially.

1. **FSR and AR overlap**

   Passing among non-ERVs in different IRs is prohibited. So, the non-ERVs in the current optimization should stop after the most downstream final position in previous optimizations. The $c$ value is increased until no FSR ends before this most downstream position.

2. **Available space vs required space**

   The ERV needs a minimum number of free cells to move along a link segment. If lane changes are performed, additional free cells are required to be able to make the maneuvers. This rule consists of ensuring that the required number of cells for ERV movement is provided by computing the available number of cells which is the difference between the total number of cells and the number of cells utilized by non-ERVs.
- **ERV speeds and desired lateral position at intersections**

As discussed in Section 5.6.1), the intersection space is not considered in the AR. Non-ERVs are not allowed to stop inside intersections and the ERV has a predefined movement to take at intersections (right, left or through) that depends on the next link in its optimal route. So, there is no need to optimize the ERV intra-link movement inside the intersection.

If the AR encloses an intersection (i.e., starts at or before an intersection and ends beyond it), a cell in the AR will either be part of the link upstream or part of the one downstream of the intersection. The most downstream cells on the link upstream of the intersection are the ones forming the stop line of the intersection. The ILP ensures that the ERV enters and exits the intersection from the appropriate lateral position depending on the intended movement inside the intersection. In addition, it is assumed that the ERV speed should decrease prior to an intersection and should be maintained inside the intersection.

If the AR does not include an intersection but ends within $(Y - 2)$ increments prior to the stop line of an intersection, the ILP makes sure that the ERV leaves the AR at the desired lateral position that depends on its upcoming movement. Otherwise, the ERV will not have enough space to make the lane changes required to enter the intersection from the appropriate lane.

Subsequently, to account for the ERV speed implications and desired ERV lateral position due to intersections, new constraints (Equations 50, 51 and 52) are added to the ILP with new parameters to be defined in this preprocessing step, as follows:

- $x_s l$ takes the value of the longitudinal index of the cells forming the stop line with respect to $AR_{start}$.
- $a^l$ takes the value of 0 if cells forming a stop line exist at the longitudinal position $x = (N + 1)i$ or between increments $i$ and $i + 1$ and 1 otherwise. Note that $x_{si}^{AR}$ is the longitudinal index of the next intersection’s stop line with respect to $AR_{start}$.

$$a^l = \begin{cases} 0 & \text{if } (N + 1)i \leq x_{si}^{AR} \leq (N + 1)i + N \\ 1 & \text{otherwise} \end{cases}$$

- $a^f$ takes the value of 0 when an ERV lateral position is desired at end of the AR due to an upcoming intersection (within $Y - 2$ increments after the AR) and 1 otherwise.

$$a^f = \begin{cases} 0 & \text{if } LL \leq x_{si}^{AR} \leq LL + (Y - 2)(N + 1)i \\ 1 & \text{otherwise} \end{cases}$$

- $y_d$ takes the value of the desired ERV lateral position that depends on the movement to be performed inside the intersection. For a link composed of 3 lanes and 1 right traversable shoulder, when a right movement is intended by the ERV at the intersection, a lateral position of 2 is desired prior that intersection. When no intersection exists within or after the AR, the constraints implying a desired lateral position will not be binding (due to $a^l$ and $a^f$ equal to 1). In that case, $y_d$ can take
any value (it is set to 1). The following desired lateral position valued change with different link composition.

\[ y_d = \begin{cases} 
2 & \text{for right movement starting } x^A_{st} \\
3 & \text{for straight movement starting } x^A_{st} \\
4 & \text{for left movement starting } x^A_{st} \\
1 & \text{no desired lateral position} 
\end{cases} \]

- **Input from output of previous optimizations**

As a sequential optimization is adopted, it is intuitive to consider the output of the previous optimization to set the initial conditions of the next ILP which are (1) ERV initial speed and lateral position, (2) previously disseminated ERV instructions, (3) previously generated ERV speed variables, and the (4) most downstream non-ERV final position in previous ARs. Note that the AR of the next ILP can overlap with a previous AR or start after the last AR and hence form a gap.

1. **ERV initial speed and lateral position**

   In case of a gap, the ERV is assumed to maintain a straight movement (same lateral position) and to increase its speed stage linearly up to \( S^{free} \) unless an intersection exists over that gap and/or in the first \((Y - 1)\) increments of the next AR. In that case, a lane change may be performed inside the gap (after exiting the last AR) based on the movement at the intersection. In addition, a speed reduction prior to the intersection is applied because it is assumed that the ERV decreases its speed before entering the intersections located in this gap. Based on these assumptions, the ERV’s initial speed and initial lateral position of the ERV are deduced for the next ILP. In the case of overlap, the ERV initial lateral position and speed are deduced by retrieving the ones that were generated from previous optimizations at the increment that coincides with the first increment in the next ILP. Nevertheless, as the ERV is travelling toward the destination, it may end up travelling at speeds that are different from the ones generated by the system. The ERV is equipped with connected vehicle technologies. So, the proposed system tracks the ERV’s actual speeds and detects any difference between the actual speed and the ones that were generated by the system. When a difference exists, the ERV speed stages at the increments that the ERV did not reach yet are adjusted to better estimate the ERV initial speed in the current longitudinal position.

2. **Previously disseminated ERV instructions**

   In case of overlap, some increments may belong to more than one AR so it important to keep the instruction consistent. In other words, the instructions that were previously disseminated to the ERV should be left unchanged in future optimizations to avoid confusion. The most downstream ERV instruction that is disseminated to the ERV is the one delineating the ERV path up to the most
downstream non-ERV final position in previous ARs. A binary parameter \((d_t_{k}^{i,y})\) takes the value of 1 if instruction \(k\) was previously disseminated to the ERV at increment \(i\) and lateral position \(y\) and 0 otherwise or when no overlap exists.

3. Previously generated ERV speed variables

In case of overlap, the ERV speed at each increment over the overlap in the next ILP should be bounded by an upper limit \((su_i)\) equal to the ERV speed environment \((s_{env}^i)\) that was previously generated at that increment, if any. The ERV speed environment is the ERV speed that takes only into consideration the presence of non-ERVs surrounding the ERV’s next movement. Setting an upper limit \((su_i)\) is needed to account for the presence of non-ERVs from previous ARs in the next ILP, in the case of overlap. The \(su_i\) takes the value of \(S_{free} + 1\) when no ERV speed environment \((s_{env}^i)\) was previously generated at that increment.

4. Most downstream non-ERV final position in previous ARs

Since no passing is allowed between non-ERVs considered in different ILP optimizations, the non-ERV with the most downstream final position and which has not yet diverged from the ERV route should be identified. This is to make sure that the non-ERVs in the next ILP are only instructed to stop at positions beyond it. The binary parameter \(np^{x,y}\) takes the value of 1 at and before the most downstream non-ERV final position in previous ARs and 0 otherwise or when no overlap exists.

5. ILP formulation

The mathematical program is as follows:

- **Objective function**

The objective function of this ILP maximizes the ERV’s speed at each increment (first component multiplied by \(\alpha_1\)) and the free space surrounding its intra-link movement (second component multiplied by \(\alpha_2\)). As previously mentioned, the \((s_{env}^i)\) is the speed environment at increment \(i\) that is only constrained by the number of non-ERVs on cells adjacent to the ERV’s movement prior to that increment (see Equations 32 through 37). If the ERV is performing a right or left lane change prior to increment \(i\), maximizing \((s_{env}^i)\) ensures that the non-ERVs are pushed away from the ERV movement as much as possible even though the ERV speed \((s_i)\) will be lower than \((s_{i-1}^i)\) due to the lane change (see Equations 40, 45 and 46). The weight factors \((\alpha_1)\) and \((\alpha_2)\) are equal to 1 since this combination is the most unbiased [1]. Adopting an objective function that is composed of the first two components only may lead to several alternative solutions. The third component (multiplied by \(\alpha_3\) that is a very small factor relative to \(\alpha_1\) and \(\alpha_2\)) is added to favor, out of these alternative solutions, the ones with the most upstream non-ERVs final positions. The reason behind this preference is to better utilize the downstream space. Due to the sequential approach adopted, non-ERVs
in the current ILP can only stop after the most downstream non-ERV final position in previous optimizations.

Maximize $z = \alpha_1 \sum_{i=2}^{L_l/(N+1)} s_i + \alpha_2 \sum_{i=2}^{L_l/(N+1)} s_{env} - \alpha_3 \sum_{x} (\sum_{y} (\sum_{j} v_{j}^{x,y} x))$ \hspace{1cm} (9)

- **Constraints**
  - A cell can be occupied by a single vehicle (ERV or non-ERV).

$$w^{x,y} + \sum_{j=1}^{l} v_{j}^{x,y} \leq 1; \forall x; \forall y \hspace{1cm} (10)$$

A non-ERV should be assigned to only one cell in its FSR.

$$\sum_{x \leq x^{'}} \sum_{y=1}^{c} v_{j}^{x,y} = 1; \forall j \hspace{1cm} (11)$$

$$\sum_{x=1}^{L_l} \sum_{y=1}^{v} v_{j}^{x,y} = 1; \forall j \hspace{1cm} (12)$$

- Passing among all non-ERVs was prohibited in [1] (Chapters 3 and 4). In this paper, this constraint is relaxed. Prohibiting passing between all non-ERVs may be too conservative in some cases. For instance, no interaction is expected when two vehicles with different lateral positions are passing each other and maintaining the same lane. When dealing with varying speeds, a vehicle $j$ initially positioned upstream of $j'$ may have a more downstream FSR than the one of $j'$. For instance, non-ERV $j$ may be travelling at a higher speed, hence requiring a lengthier minimum stopping distance. If passing is not allowed, the FSR of $j'$ has to be extended so that $j'$ is assigned to a longitudinal position equal to or higher than the one of $j$. This FSR extension increases the problem size and potentially results in higher ILP computation time. On the contrary, if $j$ is allowed to pass $j'$ (i.e., if the FSR of $j'$ is not extended), no interaction is possible as long as weaving between $j$ and $j'$ does not occur and $j$ and $j'$ are not assigned to the same lane. For instance, if $j$ is passing $j'$ and assigned to the same lane as $j'$, $j$ may do an early lane change and end up behind $j'$, which will prevent $j$ from reaching its assigned position unless it overtakes $j'$. Since the timing of non-ERV lane changes cannot be controlled, it is preferred to prevent $j$ and $j'$ to stop on the same lane when passing is occurring between this non-ERV pair to limit vehicular interactions. Equations 13 and 14 detect when two non-ERVs will pass each other to reach their generated final positions. They set the value of a binary variable ($p_{j,j'}$): it is equal to 1 if no passing occurs between $j$ and $j'$ and equal to 0 otherwise. This parameter is then used in the weaving constraints (discussed next in Equations 15 and 16) to prohibit the non-ERVs $j$ and $j'$ from weaving among each other and from stopping at the same lateral position in case passing is occurring between this pair of non-ERVs. If passing is not
happening, weaving between the pair of non-ERVs is still prohibited but non-ERVs are allowed to stop on the same lane.

\[
\sum_{y=1}^Y \sum_{x=1}^{LL} x \times v_{jy}^{x,y} \leq \sum_{y=1}^Y \sum_{x=1}^{LL} x \times v_{jy'}^{x,y} + LL(1 - p_{jj'}) \quad \forall j' > j \tag{13}
\]

\[
\sum_{y=1}^Y \sum_{x=1}^{LL} x \times v_{jy}^{x,y} + LL(p_{jj'}) \geq \sum_{y=1}^Y \sum_{x=1}^{LL} x \times v_{j'y}^{x,y} \quad \forall j' > j \tag{14}
\]

- Weaving among non-ERVs is prohibited. In addition, when passing between \( j \) and \( j' \) will occur (i.e., \( p_{jj'} = 0 \)), then \( j \) and \( j' \) are not allowed to stop on the same lateral position. For example, if passing will occur between \( j \) and \( j' \), then \( p_{jj'} = 0 \). If \( j' \) is initially to the right of \( j \) (i.e., \( y_j < y_{j'} \)), then Equation 16 applies and \( j \) cannot be assigned to the same lateral position as the one of \( j' \). It can only be assigned to a lateral position strictly greater than the one of \( j' \). If passing will not occur between \( j \) and \( j' \), then \( p_{jj'} = 1 \). If \( j' \) is initially to the left of \( j \) (i.e., \( y_j > y_{j'} \)), then Equation 15 applies and \( j \) can be assigned to the same lateral position as the one of \( j' \).

\[
\sum_{y=1}^Y \sum_{x=1}^{LL} y \times v_{jy}^{x,y} \leq \sum_{y=1}^Y \sum_{x=1}^{LL} y \times v_{jy'}^{x,y} + 1 - p_{jj'} \quad \forall j' > j \in y_j \leq y_{j'} \tag{15}
\]

\[
\sum_{y=1}^Y \sum_{x=1}^{LL} y \times v_{jy}^{x,y} \geq \sum_{y=1}^Y \sum_{x=1}^{LL} y \times v_{jy'}^{x,y} + 1 - p_{jj'} \quad \forall j' > j \in y_j > y_{j'} \tag{16}
\]

- A minimum of one cell should be provided for the ERV at each longitudinal position.

\[
\sum_{y=1}^Y \sum_{j=1}^J v_{jy}^{x,y} \leq Y - 1; \quad \forall x \tag{17}
\]

- The ERV can receive only one instruction at each increment.

\[
\sum_{y=1}^Y \sum_{k=1}^3 d_{ky}^{i,y} = 1; \quad \forall i = 1, ..., LL/(N + 1) - 1 \tag{18}
\]

- Go right and go left instructions cannot be generated at the rightmost lane and leftmost lanes, respectively.

\[
d_1^{i,y} = 0; \quad \forall i = 1, ..., LL/(N + 1) - 1 \tag{19}
\]

\[
d_3^{i,y} = 0; \quad \forall i = 1, ..., LL/(N + 1) - 1 \tag{20}
\]

- If the current AR overlaps with previous ARs, the previously disseminated ERV instruction should be left unchanged, as discussed in Section 5. 6. 3).

\[
d_k^{i,y} \geq d_{-k}^{i,y}; \quad \forall k; \quad \forall i = 1, ..., LL/(N + 1) - 1; \quad \forall y \tag{21}
\]

- The ERV instruction and ERV assignment variables should be interconnected to ensure the continuity of the longitudinal and lateral ERV movement. When a given instruction is generated at an increment, the corresponding cells should be added to the ERV intra-link path.

\[
\sum_{y=1}^Y w_{x,y}^{x,y} = 1; \quad \forall x \tag{22}
\]
\[
\begin{align*}
  d^{i,y}_1 & \leq \frac{w^{i(N+1),y} + w^{i(N+1)+t,y-1}}{2}; \quad \forall i = 1, \ldots, \frac{LL_{N+1}}{N+1} - 1; \forall t = 1,2, \ldots, N+1; \forall y > 1 \\
  d^{i,y}_1 & \geq w^{i(N+1),y} + w^{i(N+1)+t,y-1} - 1; \quad \forall i = 1, \ldots, \frac{LL_{N+1}}{N+1} - 1; \forall t = 1,2, \ldots, N+1; \forall y > 1 \\
  d^{i,y}_2 & \leq \frac{w^{i(N+1),y} + w^{i(N+1)+t,y+1}}{2}; \quad \forall i = 1, \ldots, \frac{LL_{N+1}}{N+1} - 1; \forall t = 1,2, \ldots, N+1; \forall y < Y \\
  d^{i,y}_2 & \geq w^{i(N+1),y} + w^{i(N+1)+t,y+1} - 1; \quad \forall i = 1, \ldots, \frac{LL_{N+1}}{N+1} - 1; \forall t = 1,2, \ldots, N+1; \forall y < Y
\end{align*}
\]

- The ERV instruction and non-ERV assignment variables should also be interconnected to ensure that the non-ERVs are not positioned on cells that are part of the ERV path or on the additional cells needed to perform right or left instructions.

\[
\begin{align*}
  \sum_{x=t'} \sum_{j=1}^{l} v^{x,y}_j & \leq (N + 1) \left( 1 - d^{i,y}_2 \right); \\
  \forall i = 1, \ldots, \frac{LL_{N+1}}{N+1} - 1; t' = (N + 1)i + 1; t'' = (N + 1)i + (N + 1) \\
  \sum_{x=t'} \sum_{j=1}^{l} v^{x,y}_j + \sum_{x=t} \sum_{j=1}^{l} v^{x,y-1}_j & \leq (2N + 1) \left( 1 - d^{i,y}_3 \right); \forall y > 1; \\
  \forall i = 1, \ldots, \frac{LL_{N+1}}{N+1} - 1; t' = (N + 1)i + 1; t'' = (N + 1)i + (N + 1); t''' = (N + 1)i + N \\
  \sum_{x=t'} \sum_{j=1}^{l} v^{x,y}_j + \sum_{x=t} \sum_{j=1}^{l} v^{x,y+1}_j & \leq (2N + 1) \left( 1 - d^{i,y}_3 \right); \forall y < Y; \\
  \forall i = 1, \ldots, \frac{LL_{N+1}}{N+1} - 1; t' = (N + 1)i + 1; t'' = (N + 1)i + (N + 1); t''' = (N + 1)i + N
\end{align*}
\]

- The ERV speed environment is constrained by the presence of non-ERVs around the ERV’s next movement only. For example, if the next movement does not have any non-ERVs stopped on its adjacent cells, then based on the surroundings, the ERV’s speed stage can increase by 1. In case of overlap, non-ERVs can only be assigned beyond the most downstream non-ERV’s final position in previous ARs. Thus, the speed environment variables along the overlap in the current AR consider the presence of non-ERVs in previous ARs. This is performed by limiting the ERV speed environment in the current ILP to speed upper limits determined in Section 5.6.3).

\[
\begin{align*}
  s_{env}^{i+1} & \leq s^i + 1 - \sum_{j=1}^{l} v^{(N+1)i+t,y-1}_j - \sum_{j=1}^{l} v^{(N+1)i+t,y+1}_j + S^{free} \left( 1 - w^{(N+1)i+t,y} \right); \\
  \forall i = 1, \ldots, \frac{LL_{N+1}}{N+1} - 2; \forall t = 0,1, \ldots, N + 2; \forall y = 2,3, \ldots, Y - 1 \\
  \frac{LL}{S_{env}^{N+1}} & \leq \frac{LL}{S_{env}^{N+1}} - 1 + \sum_{j=1}^{j} v^{LL-(N+1)+t,y-1}_j - \sum_{j=1}^{j} v^{LL-(N+1)+t,y+1}_j + S^{free} \left( 1 - w^{LL-(N+1)+t,y} \right) \\
  \forall t = 0,1, \ldots, N + 1; \forall y = 2,3, \ldots, Y - 1
\end{align*}
\]
\[ s_{env}^{i+1} \leq s_{env}^i + 1 - \sum_{j=1}^I v_j^{(N+1)i+t,2} + S_{free}(1 - w^{(N+1)i+t,1}); \]
\[ \forall i = 1, ..., \frac{LL}{N+1} - 2; \forall t = 0,1, ..., N + 2 \quad (34) \]
\[ s_{env}^{i+1} \leq s_{N+1}^{i+1} + 1 - \sum_{j=1}^I v_j^{LL-(N+1)+t,2} + S_{free}(1 - w^{LL-(N+1)+t,1}); \forall t = 0,1, ..., N + 1 \quad (35) \]
\[ s_{env}^{i+1} \leq s_{i+1} + \sum_{j=1}^I v_j^{(N+1)i+t,Y-1} + S_{free}(1 - w^{(N+1)i+t,Y}); \]
\[ \forall i = 1, ..., \frac{LL}{N+1} - 2; \forall t = 0,1, ..., N + 2 \quad (36) \]
\[ s_{env}^{i+1} \leq s_{LL}^{i+1} + 1 - \sum_{j=1}^I v_j^{LL-(N+1)+t,Y-1} + S_{free}(1 - w^{LL-(N+1)+t,Y}); \forall t = 0,1, ..., N + 1 \quad (37) \]
\[ s_{env}^{i+1} \leq s_i + 1; \forall i = 1, ..., \frac{LL}{N+1} - 1 \quad (38) \]

- The ERV temporary speed variable at an increment is limited by the ERV speed environment at that increment, by the ERV instruction at the previous increment and the presence of an intersection. In addition, and as discussed in 5. 6. 3), the ERV temporary speed variable at increment \( i \) decreases by one if increment \( i \) is at or before an intersection (Equation 41 and 42). The ERV temporary speed is maintained until the ERV exits the intersection (Equation 43).

\[ s_{temp}^{i+1} \leq s_{env}^i; \forall i = 1, ..., \frac{LL}{N+1} - 1 \quad (39) \]
\[ s_{temp}^{i+1} \leq s_i + 2 \sum_{j=1}^Y d_2^{j,Y} - 1; \forall i = 1, ..., \frac{LL}{N+1} - 1 \quad (40) \]
\[ s_{temp}^{i+1} \leq s_i + 2 \alpha^{i+1} - 1; \forall i = 1, ..., \frac{LL}{N+1} - 2 \quad (41) \]
\[ s_{temp}^{i+1} \leq s_{LL/(N+1)-1} + 2\alpha^f - 1 \quad (42) \]
\[ s_{temp}^{i+1} \leq s_i + \alpha^i; \forall i = 1, ..., \frac{LL}{N+1} - 1 \quad (43) \]

- The ERV speed at increment \( i \) should be equal to the maximum of \( S_{min} \) and \( s_{temp}^i \) and cannot increase beyond the maximum allowable speed \( S_{free} \).

\[ s^{i+1} \geq S_{min}; \forall i = 1, ..., \frac{LL}{N+1} - 1 \quad (44) \]
\[ s^{i+1} \geq s_{temp}^{i+1}; \forall i = 1, ..., \frac{LL}{N+1} - 1 \quad (45) \]
\[ s^{i+1} \leq S_{free}(1 - v^{i+1}) + s_{temp}^{i+1}; \forall i = 1, ..., \frac{LL}{N+1} - 1 \quad (46) \]
\[ s^{i+1} \leq (S_{free} - S_{min}) v^{i+1} + S_{min}; \forall i = 1, ..., \frac{LL}{N+1} - 1 \quad (47) \]
\[ s^{i+1} \leq S_{free}; \forall i = 1, ..., \frac{LL}{N+1} - 1 \quad (48) \]

- Estimated unconnected non-ERVs are assumed to follow the behavior of their respective leader. This constraint virtually reserves a cell behind the leader of an estimated
unconnected non-ERV for use by the actually present unconnected non-ERVs. This constraint is only binding for unconnected non-ERVs (when $t_y=0$).

$$\sum_{x'=1}^{x} v^j_{x'} + t_y \geq v_{x}^{x'}; \quad \forall j \in LL; \forall y$$ (49)

- A final ERV lateral position should be imposed if an intersection exists at the end of the AR or within $(Y - 2)$ increments of its end.

$$w^{(LL,y_d)} + \alpha^f \geq 1 \quad (50)$$

- If the stop line of an intersection is inside the AR, then, a lateral position is to be imposed before and after the intersection depending on the future movement of the ERV inside the intersection. $y_d$ is the desired lateral position that depends on the ERV’s movement inside the intersection. Equation 51 sets the desired lateral position before entering the intersection and Equation 52 sets the desired lateral position when exiting the intersection (i.e. entering the downstream link). It is assumed that the same lateral position is maintained when exiting the intersection. This is why a “go straight” instruction will be indirectly generated at the increment located at or before the stop line. This “go straight” instruction will not be disseminated to the ERV to avoid confusion.

$$w^{((N+1)(i),y_d)} + \alpha^i \geq 1; \quad \forall i = 1,\ldots,\frac{LL}{N+1} - 1 \quad (51)$$

$$w^{((N+1)(i+1),y_d)} + \alpha^i \geq 1; \quad \forall i = 1,\ldots,\frac{LL}{N+1} - 1 \quad (52)$$

5.7. Experimental Analysis

The proposed system is modeled in a simulation environment to assess its impact on traffic operations in an urban transportation network with signalized intersections. A 4 by 4 grid transportation network shown in

Figure 5- 8) is modeled in NetLogo, an agent-based modeling language and environment. NetLogo is a widely used and open source tool that has already proved to be able to support microscopic traffic simulations [28]. The ability to implement the proposed complex system while closely visualizing the intra-link behavior of each vehicle motivated us to use this tool for the experimental analysis [28]. Agent-based models are practical when simulating simultaneous operations and interactions of multiple agents. In our case, each vehicle (ERV or non-ERV) is a turtle that moves around in the NetLogo world. It reacts to its surroundings while aiming to follow any real-time message it may receive if equipped with connected vehicles technologies. Patches are the agents in NetLogo that form the ground over which the vehicles move [29]. This discrete aspect of the NetLogo world (space) fits the proposed system which also discretizes transportation links into identical cells with a size equal to that of a vehicle plus buffer. Using APIs, the model in NetLogo is invoked and controlled by a master program running on the java virtual machine. This java program activates the hyperstar routing algorithm, when the ERV route generation module is triggered.
The CAM, coded in NetLogo, periodically scans and filters the non-ERV of interest. When a critical group of non-ERV is detected, the sequential ILP optimization module, coded in a separate java program, is executed. The latter preprocesses the connected vehicles data from NetLogo and initiates the ILP that is coded in AMPL and solved using CPLEX. All tests are run on a MacBook Pro computer with 3.1 GHz Intel Core i5 and 8 GB 2133 MHz LPDDR3 memory.

Due to the presence of signalized intersections in the system, emergency preemption should be implemented. This concept, widely deployed at signalized intersections in the United States [30], consists of granting priority to the emergency response by requesting then displaying a green signal to its approach upon detection while prohibiting vehicles on the conflicting approaches from entering the intersection. Preemption reduces the ERV’s travel time as it eliminates the risk of having an ERV waiting at a red signal. It is also activated prior to the ERV's arrival to discharge the downstream queue [9], and this is required to create gaps between vehicles to give them the opportunity to move and pull over [11]. The focus of this work is on the intra-link ERV movement. In real-life implementation of this system, any traditional or advanced emergency preemption technique can be used at signalized intersections. In this experimental analysis, a dynamic ERV detection distance is adopted for emergency preemption and is obtained based on [11, 31]. The ERV detection distance depends on the time needed to switch the green indication to the ERV’s approach, the average queue length on the approach, and the ERV operating speed. Sensing technologies such as pavement inductive loops or microwave radars [11, 31] are assumed present for the estimation of the queue length. So, the dynamic detection distance is assumed to be insensitive to the market penetration level. This will allow a better assessment of the benefits due to the proposed intra-link emergency assistance since these benefits are isolated from the ones due to adequate preemption activation, when compared to situations with zero market penetrations.

Figure 5-8: Network representation showing the link and node labels
Signals timing plans can be fixed (pre-timed), actuated, or adaptive [11]. Since emergency preemption is implemented at signalized intersections, it is expected to interrupt the cycle when activated, so the differences between the various types of signal timing plans can be considered minimal. In addition, since the focus of this system is the intra-link traffic operations and for simplicity, a pre-timed plan is adopted at intersections and maintained for all tests to ensure consistent comparison of the results.

Each connected non-ERV that is detected by the CAM and whose data is sent to the SIOM, receives its instruction message when the corresponding ILP’s computation time has elapsed. It is assumed that the connected vehicle needs 2.5 seconds of reaction time [32] before starting to brake and to move towards its assigned position. On the other hand, an unconnected vehicle which cannot be identified by the CAM nor provided with a warning/instruction message, only detects the approaching ERV when located within the latter’s siren range, which is set to 555 ft [33]. Upon detecting the ERV and after 2.5 seconds of reaction time, if the non-ERV is following another non-ERV, it is assumed to follow the preceding vehicle’s behavior. Otherwise, it is assumed that the unconnected vehicle starts braking and aiming for an empty gap on its nearest edge. If no empty gap is available on the nearest edge and if the ERV is directly behind the non-ERV then it is assumed that the unconnected non-ERV tries to move to the other edge, as the ERV location is now visible. If both sides are blocked and the ERV is directly behind, the unconnected vehicle will start accelerating while continuously trying to shift lanes. Similarly, it is assumed that a connected non-ERV that is still trying to reach its assigned position or that fails to do so possibly due to insufficient time resulting from higher than expected ILP computation time, drops the instruction and acts as an unconnected non-ERV if the ERV is directly behind.

The proposed system is tested under three v/c ratios of 0.75, 1 and 1.25, and different levels of market penetrations ranging between 0 and 100% at 20% increments. For the experiments conducted here, each link is composed of 3 lanes and 1 traversable shoulder. For each combination of market penetration level and congestion level, the results are compared to the case with no system (i.e., at 0% market penetration level). The performance measures used to assess the benefits of the proposed system are the ERV travel time and vehicular interactions (ERV/non-ERV and non-ERV/non-ERV). To quantify the impact of the proposed system, the average time during which non-ERVs where affected by the approaching ERV is recorded (discussed further in Section 5. 8. 3). In other words, the proposed system may result in notifying connected non-ERVs about an approaching ERV early in time, increasing the time during which they are affected by the ERV passage.

5. 8. Discussion of results

The use of NetLogo for the visualization of the proposed system is certainly valuable to observe and validate the system’s functionality. In addition, it allows a timely assessment of the impacts of the
proposed system. The proposed system is developed to improve the ERV’s travel time, its safety and that of its surrounding vehicles. By incrementally providing the ERV with maneuvering assistance and downstream vehicles with early warning/recommendation messages, the system is expected to assist the ERV as well as non-ERVs. Yet, some of assumptions made may not always be satisfied, hence it is important to emulate real-world conditions in a simulation environment to understand the true benefits/impacts the system. For example, the minimum stopping distance computed for each non-ERV in Section 5.6.3), to ensure that the non-ERV is assigned to a position it can reach comfortably, includes the distance travelled during the expected ILP computation time. If the actual computation time is higher than the expected one used in the MSD calculations, then the non-ERV will start braking later than expected, hence reaching the assigned final position on time is not guaranteed. As discussed in Section 5.7., the system is executed on an urban transportation network with signalized intersections. A Poisson distribution is used to model traffic on each link in the network with average flow rates for each v/c ratio based on [34]. After the seed period, traffic will stop being generated as vehicles leaving the network (i.e. NetLogo world) will reenter it from the opposite edge because the NetLogo world topology wraps vertically and horizontally (i.e. opposite sides are actually connected). This will emulate the presence of other intersections outside the network we modeled. For each combination of v/c ratio and market penetration level, 5 runs were initially conducted. Then, based on the average ERV travel time and standard deviation, the minimum sample size for each combination is identified using a 90% confidence interval and a 5 second desired margin of error. Accordingly, more runs are executed until minimum sample size is satisfied. The next subsection 5.8.1) discusses the benefits obtained in terms of ERV travel time for the different combinations of v/c ratios and market penetration levels while subsection 5.8.2) presents the benefits in terms of vehicular interactions. Finally, subsection 5.8.3) clarifies the impact that the system has on non-ERVs drivers and further investigates the trade-off between this impact and the observed benefits.

5.8.1) ERV travel time benefits

The reduction in ERV travel time is the main indication of the effectiveness of an emergency assistance system. Figures V-9, V-10 and V-11 show the variation in percent reduction in ERV travel time with market penetration levels, when compared to the no-system scenario with 0% market penetration, for v/c ratios of 0.75, 1 and 1.25 respectively. The proposed system succeeds in improving ERV travel time at all v/c ratios when the market penetration level is equal to or higher than 40%. As shown in Figure V-10 for a v/c ratio equal to 1, an increase in ERV travel time is noted, which makes the proposed system not recommended at 20% market penetration although significant reductions in vehicular interactions potentially result (discussed in Section 5.8.2)). For all v/c ratios, the observed percent reduction in ERV travel time increases as the market penetration level increases. This increasing trend is expected. With a
higher proportion of connected non-ERVs on the link, a faster passage for the ERV is more achievable. With partial market penetration, higher risks of interactions between the ERV and non-ERVs exist (discussed in Section 5. 8. 2), hence negatively affecting the ERV’s movement. Unconnected non-ERVs stopped on the nearest edge may force an ERV that was travelling on the edge to deviate to another free lane, resulting in a speed decrease. In addition, unconnected non-ERVs may occupy a position at which a connected non-ERV is instructed to stop. Hence, this connected non-ERV may end up impacting the ERV movement by stopping in a cell adjacent to the ERV’s path. At higher market penetration, the impact of unconnected vehicles on the ERV travel is minimized. At 100% market penetration, the highest percent reductions in ERV travel time are observed since all non-ERVs are moving to their assigned positions that are not at risk of being occupied by other unconnected non-ERVs. It is important to note that this system may not offer tremendous benefits in terms of ERV travel time (range between 1% and 11%) as the ILP is generating the ERV passage while setting very restrictive rules about the weaving interactions between non-ERVs. For instance, the ERV optimal movement may include a right/or left lane change, that results in a speed decrease, just to avoid having non-ERVs stopped adjacent to its movement. If such lane change is perceived as unnecessary (more weight can be attributed to the first component in the objective function compared to the second one, i.e., $\alpha_1 > \alpha_2$ can be used). In addition, the system may instruct non-ERVs to stop at locations adjacent to the ERV lane if this is the only position that would satisfy the no-weaving rule. If this is considered as too conservative, the relaxation of the corresponding constraints in the ILP can be made. The next subsection investigates the benefits of this system in terms of vehicular interactions.

![Figure 5-9: Percent reduction in ERV travel time for each market penetration level at a v/c ratio equal to 0.75](image)

Figure 5-9: Percent reduction in ERV travel time for each market penetration level at a v/c ratio equal to 0.75
5.8.2 Vehicular interactions

Limiting vehicular interactions among non-ERVs and between the ERV and non-ERVs is a major contribution offered by this system. Making sure that the ERV can move forward and safely is paramount but ensuring that non-ERVs are also safe is what makes this system more valuable. Each non-ERV is assigned to a position that is at or beyond its minimum stopping distance, while limiting weaving and passing with others. Interactions among non-ERVs is assessed by counting risky maneuvers between non-ERVs. A risky interaction between two non-ERVs is detected when both were aiming towards the same lane and one non-ERV had to pass the other. On the other hand, the ILP generates recommendation messages that try to direct the non-ERVs away as much as possible from the ERV’s movement (by maximizing the second component in the objective function). To assess the amount of ERV/non-ERV interactions, the number of non-ERVs adjacent to the ERV’s movement are counted for each scenario.
Figure 5-12 shows the average percent reduction for each of the non-ERV/non-ERV and ERV/non-ERV interactions for different market penetration levels. A reduction is observed at all market penetration levels for both types of vehicular interactions along with an increasing trend as the market penetration level is higher, hence confirming that this system enhances the safety of vehicles. Higher reductions in non-ERV/non-ERV interactions are observed when compared to the one between ERV and non-ERVs. This is expected because as mentioned before, a feasible solution generated by the ILP may assign the non-ERVs to positions that are adjacent to the ERV path only to satisfy the no-weaving constraint.

![Percent reduction in vehicular interactions for different market penetrations](image)

**Figure 5-12: Percent reduction in vehicular interactions for different market penetration levels**

5. 8. 3) Impact of the proposed system and limitations

From the non-ERVs’ perspective, an ideal emergency assistance system notifies them only when it is needed, in a way to grant them enough time to move and create the optimal passage for the ERV while most importantly minimizing the time they have to wait for the ERV at their final position. A system sending unnecessary recommendation messages to non-ERVs and/or requesting them to wait for a very long duration before the ERV’s passage is not effective and will definitely not be embraced. By considering the non-ERVs within a detection range in the CAM (Section 5. 5. 1)) and then filtering out the non-ERVs that are diverging from the ERV’s route before its arrival (Section 5. 5. 2)), the system identified the non-ERVs of interest which should be provided with an instruction message. To assess the impact of the proposed system on non-ERVs, the average time during which non-ERVs were affected by the ERV presence is recorded at each market penetration level. This duration starts upon receiving the instruction message and ends when the ERV becomes downstream. Note that non-ERVs’ recovery after the ERV’s
passage if not investigated in this paper and will be included in future works. In Table 5-3, the average percent increase in impact on non-ERVs (in terms of duration affected by the ERV passage) when compared to the case at 0% market penetration, is shown for each market penetration level. The system did not offer a reduction in ERV travel time at a 20% market penetration which means that the system was ineffective, hence the lowest impact on the non-ERVs. At market penetrations higher than 20%, the average percent increase in impact on non-ERVs decreases as the market penetration become higher. With a decreasing proportion of unconnected non-ERVs on the link, the connected non-ERVs are capable of reaching their final positions faster.

Table 5-3: Percent increase in impact (in terms of duration affected by the ERV passage) on non-ERVs at each market penetration level

<table>
<thead>
<tr>
<th>MP (%)</th>
<th>Percent increase in impact on non-ERVs</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>6.47%</td>
</tr>
<tr>
<td>40</td>
<td>9.47%</td>
</tr>
<tr>
<td>60</td>
<td>8.28%</td>
</tr>
<tr>
<td>80</td>
<td>7.30%</td>
</tr>
<tr>
<td>100</td>
<td>7.30%</td>
</tr>
</tbody>
</table>

The detection distance ($\Delta d_{det}$) includes a factor of safety ($FS$) primarily to account for the additional time required by non-ERVs to change lanes. In all the tests executed above, $FS$ is set to 1.25. An increase in $FS$ will directly result in higher detection distances, which should potentially lead to better ERV travel times because non-ERVs have more time to react and reach their assigned positions. In other words, due to possible inaccuracies of the heuristic approach adopted in the CAM (such as the use of an expected ILP computation time), the estimated duration that a non-ERV needs to reach its final position may be underestimated. To confirm this trade-off, a set of additional tests are executed with a factor of safety equal to 1.5 at 100% market penetration. The average reduction in ERV travel time increased from 9.09% (with $FS=1.25$) to 15.40% (with $FS=1.5$) while the average increase in non-ERV impact increased from 7.30% to 11.60%. These results are expected. Increasing the $FS$ used in the detection distance calculation leads to more benefits in terms of ERV travel time associated with more non-ERV impacts. With higher factors of safety, non-ERVs are waiting for a lengthier period of time at their final positions.

5.9. Conclusion

Sirens and strobe lights have been in use for decades to prioritize ERVs’ movements along streets. Yet, these traditional systems are not efficient when it comes to preventing confusion, as downstream traffic may still be unsure about how to properly react. With the emerging capabilities of ITS technologies, more advanced emergency assistance systems can be developed. This paper introduces a system that facilitate
the movement of an ERV in an urban transportation network from an origin to a destination by also focusing on link level support, an aspect of emergency assistance that has not received enough attention previously. The system reduces confusion by sending maneuver recommendations to downstream vehicles. It accounts for market penetration and is adapted to urban transportation networks with intersections. The system consists of three modules. The first module is the ERV route generation module that determines the ERV route from origin to destination. This module relies on a vehicle routing algorithm. The time-dependent hyperstar routing algorithm is used in this paper as it has been identified as a good fit for ERV routing due to its dynamic aspect. The second module is a criticality analysis module (CAM) that identifies, based on a heuristic approach, the downstream non-ERVs that should receive an instruction message. As the ERV is moving towards its destination, new non-ERVs are detected by the second module (CAM). The information of these connected non-ERVs is sent to the third module called the sequential ILP optimization. The latter estimates the presence of unconnected non-ERVs based on a distance criterion, executes a set of preprocessing steps then solves a mathematical program. The ILP generates the optimal ERV intra-link movement along a downstream range on the ERV’s route, along with the instruction messages to the non-ERVs.

To evaluate the benefits and impacts of the proposed system a grid transportation network is modeled in NetLogo, an agent-based modeling environment. Tests were executed for different combinations of congestion and market penetration levels. Results shows reduction in ERV travel time (with an average of 9.09% at 100% market penetration), with a minimum recommended market penetration of 40%. More significant reductions are observed in terms of vehicular interactions (with an average of 35.46% and 81.38% for ERV/non-ERV and non-ERV/non-ERV interactions respectively, at 100% market penetration), as the ILP maximizes the gap around the ERV movement and prohibits weaving between non-ERVs. Results confirms that this novel system that is bringing a new aspect of ERV support to light does not only prioritize the ERV’s movement by providing it a safer and clearer passage but also ensures safer movements of downstream vehicles when compared to the currently deployed techniques for ERV assistance.

Further extensions, such as the accommodation of multiple emergencies, will be addressed in future works. To reduce the impact of the ERV’s passage on downstream traffic, driver’s route information will be considered when available. In addition, technique for efficient and safe recovery after the ERV’s passage will be developed to assist vehicles resuming their movement. Finally, the implications and countermeasures of imperfect communications will be addressed to prepare this system for short-deployment.
Appendix D

The identification of the $\Delta d_{det}$ is based on a time-space diagram. An ERV and non-ERV are conflicting when their trajectories meet (see conflict point of original non-ERV trajectory and ERV trajectory in Figure D-1). If the non-ERV is instructed to brake to a stop at a specific position, the ERV and non-ERV will meet earlier in time (see conflict point of the non-ERV trajectory and ERV trajectory in Figure D-1). In all of the following figures, the ERV and non-ERVs are travelling in the same direction.

![Diagram showing time-space relationship between ERV and non-ERV trajectories]

Figure D-1: Sample time-space diagram of conflicting ERV and non-ERV

The $\Delta t_{clr}$ is the clearing time interval that a non-ERV needs from the time it receives the instruction until the time it stops. The $\Delta t_{clr}$ includes the reaction and braking durations. The $\Delta d_{clr}$ is the corresponding clearing distance travelled during the reaction and braking time.

$$\Delta d_{clr} = t_r v_v + \frac{v_v^2}{2g}$$

$$\Delta t_{clr} = t_r + \frac{v_v}{g}$$

Where: $(v_v)$ is the current non-ERV speed (ft/sec)
$(t_r)$ is the non-ERV’s reaction time (sec)
$(g)$ is the non-ERV’s deceleration (ft/sec$^2$)
The difference in time between the original and new conflict points ($\Delta t^*$) (non-ERV decelerates for ERV vs. non-ERV continues original movement) is:

$$\Delta t^* = \frac{\Delta t_{clr} v_e - \Delta d_{clr}}{(v_e - v_p)} = \frac{v_p^2 / 2g}{(v_e - v_p)}$$

Where: ($v_e$) is the current non-ERV speed (ft/sec)

($v_p$) is the current ERV speed (ft/sec)

($\Delta d_{clr}$) is the distance for reaction and braking (ft)

($\Delta t_{clr}$) is the reaction and braking time interval (sec)

($g$) is the non-ERV’s deceleration (ft/sec$^2$)

The minimum detection distance needed is shown below. A factor of safety (greater if equal to 1) is added to account for additional time needed to make lane changes.

$$\Delta d_{det} = (v_e - v_p) * (\Delta t_{cc} + \Delta t_{clr} + \Delta t^*)$$

$$\Delta d_{det} = (v_e - v_p) * (\Delta t_{cc} + t_r + \frac{v_p}{g} + \frac{v_p^2 / 2g}{(v_e - v_p)})$$

$$\Delta d_{det} = FS[(v_e - v_p) * (\Delta t_{cc} + t_r + \frac{v_p}{g} + \frac{v_p^2}{2g})]$$

Where: ($v_e$) is the current non-ERV speed (ft/sec)

($v_p$) is the current ERV speed (ft/sec)

($t_r$) is the non-ERV’s reaction time (sec)

($\Delta d_{clr}$) is the distance for reaction and braking (ft)

($\Delta t_{clr}$) is the reaction and braking time interval (sec)

($g$) is the non-ERV’s deceleration (ft/sec$^2$)

($\Delta t_{cc}$) is the communication and computation time interval (sec)

($\Delta t^*$) is the difference in time between the original and new conflict points

(FS) is a factor of safety to account for additional time needed for lane changes

$\Delta d_{det}$ depends on:

- Current speed of the ERV
- Average speed of the downstream non-ERVs within a distance of X ft where X is the expected maximum detection range (found using a look up table for different parameters).
- Expected computation and communication time
- Non-ERV reaction time
- Non-ERV deceleration rate
Based on the time-space diagram of Figure D-2, the minimum value for \( \varepsilon \) is determined by identifying a non-ERV that will become critical after \((\Delta t)\) which is the time-step of CAM:

\[
\Delta d_{det} + \varepsilon + v_0 \Delta t = \Delta d_{det} + v_0 \Delta t
\]

\[
\varepsilon = v_e \Delta t - v_0 \Delta t = \Delta t (v_e - v_0)
\]

\[
\varepsilon = \Delta t (v_e - v_0)
\]

Where: \((v_e)\) is the average non-ERVs speed (ft/sec)

\((v_0)\) is the current ERV speed (ft/sec)

\((\Delta t)\) is time step at which CAM is executed (sec)

According to Figure D-3, if non-ERV 1 is at a distance of \(\Delta d_{det} + \varepsilon\) downstream of the ERV, other non-ERVs are expected to become critical during the \(\Delta t_{cc}\) of the optimization triggered by non-ERV 1. These non-ERVs are the ones at a distance of \(\Delta d_{det} + \varepsilon + \Delta d_{add1}\) where \(\Delta d_{add1}\) is an additional distance determined below. In Figure D-3, non-ERVs 2 and 3 have been added to the same ILP optimization that non-ERV 1 will undergo. Non-ERV 2 is expected to be critical and at a distance of \(\Delta d_{det}\) downstream of the ERV during \(\Delta t_{cc}\) of the optimization triggered by non-ERV 1. Non-ERV 3 is expected to be at a distance of \(\Delta d_{det}\) downstream of the ERV within the time step \(\Delta t\) that follows the \(\Delta t_{cc}\). This is why it is also considered critical when non-ERV 1 is detected and this is why the additional distance \(\Delta d_{add1}\) is an additional distance measured beyond \(\Delta d_{det} + \varepsilon\). Non-ERV 4 may not be added to ILP optimization as it
becomes critical and a distance of \((\Delta d_{det} + \varepsilon)\) downstream the ERV after the system generates the instructions of non-ERVs 1, 2 and 3; a new optimization can be executed for non-ERV 4. 

\(\Delta d_{add1}\) is computed as follows and only when at least one non-ERV is detected at a distance of \((\Delta d_{det} + \varepsilon)\) beyond the ERV.

\[
\begin{align*}
\Delta d_{det} + \varepsilon + \Delta d_{add1} + v_y\Delta t_{cc} &= \Delta d_{det} + \varepsilon + v_e\Delta t_{cc} \\
\Delta d_{add1} + v_y\Delta t_{cc} &= v_e\Delta t_{cc} \\
\Delta d_{add1} &= v_e\Delta t_{cc} - v_y\Delta t_{cc} \\
\Delta d_{add1} &= \Delta t_{cc}(v_e - v_y)
\end{align*}
\]

Where: \((v_y)\) is the average non-ERVs speed (ft/sec) 

\((v_e)\) is the current ERV speed (ft/sec) 

\((\Delta t_{cc})\) is the communication and computation time interval (sec)

As discussed in Section 5.5.1), new non-ERVs are added to the group of critical ones if their estimated waiting time at a final position does not exceed a maximum tolerated duration \((\Delta t_{wait})\).

According to Figure D- 4, if \(\Delta t_{wait}\) is the maximum allowable time duration that a non-ERV tolerates waiting at its final instructed position before ERV passage, then the vehicles that can be added to the critical non-ERVs group are the ones at a distance of \((\Delta d_{add2})\) beyond \((\Delta d_{det} + \varepsilon)\). Note that the selection of the \(\Delta t_{wait}\) is subjective.
\[ \Delta d_{add2} = v_\text{e} \Delta t_{\text{wait}} \]

Where: \((v_\text{e})\) is the predicted ERV speed (ft/sec)

\((\Delta t_{\text{wait}})\) is the maximum allowable waiting duration (sec)

As discussed in Section 5.5.3, the system benefits from traffic control prior to the ERV’s station to prohibit new non-ERVs from entering the first link in the ERV’s route for a short duration of time \((\Delta t_{\text{ne}})\) before the ERV’s desired departure. Figure D-5 shows \(\Delta t_{\text{ne}}\) which is the time duration before the ERV’s desired departure time during which the upstream intersection prohibits the entry of new vehicles into the ERV’s first link. \(\Delta t_{\text{ne}}\) is computed as follows, where \(v_\text{r}\) is the average speed (ft/sec) of the non-ERVs downstream the ERV, \(\Delta d_{\text{det}}\) and \(\varepsilon\) (ft) are the minimum and additional detection distances respectively and computed as follows:

\[ \Delta t_{\text{ne}} = \frac{\Delta d_{\text{det}} + \varepsilon}{v_\text{r}} \]
Figure D-5: Prohibit entry of non-ERVs $\Delta t_{ne}$ prior $t_{dd}$

- $t_{dd}$: Desired ERV departure time
- $\Delta t_{ne}$: Time interval before the $t_{dd}$ during which entry to the first link in ERV route is prohibited
- $\Delta t_{det}$: Minimum detection distance
- $\epsilon$: Additional detection distance to account for frequency of the CAM execution
Reference


Chapter 6. CONCLUSION

In conclusion, this dissertation proposes an advanced emergency response assistance system that improves the ERV’s intra-link movement from origin to destination in an urban transportation network, an aspect of emergency assistance that has not received enough research attention previously.

People’s lives often depend on the emergency response services. Unfortunately, their operations on the roads are characterized by high risks of vehicle crashes and significant response time. Currently deployed measures for emergency support primarily include preemption at signalized intersections. It consists of granting the ERV’s approach an early or extended green signal to ensure that the ERV does not wait and is not stuck in a queue prior to entering the intersection. Assistance along transportation links is provided through the use of sirens and strobe lights. Unfortunately, these traditional warning systems are not fully effective as downstream vehicles are often unsure about how to react [1]. Further emergency support along transportation links is needed. Advanced emergency warning systems were developed to provide downstream traffic with early notification and relevant information to help them react properly and in a timely manner [2-4]. Nevertheless, these systems do not guarantee that drivers will act cooperatively. Furthermore, as decision-making is transitioning from human-based to computer-based, systems that generate the optimal actions to be taken by vehicles downstream an ERV are important.

Chapter 3 introduces the core of the proposed system which is a mathematical program (ILP) that generates the fastest ERV intra-link movement along a transportation link segment using the capabilities of the connected vehicle environment. By maximizing the ERV’s speed and free space around the ERV’s movement, the ILP generates ERV maneuvering instructions as well as a unique warning message that also includes a recommended final position for each connected non-ERV. The non-ERVs positions can be reached comfortably and are generated in a way to limit vehicular interactions, hence maintaining the safety of the ERV as well as non-ERVs. Results from a sensitivity analysis showed that the ILP computation time mainly increases with the number of non-ERVs. A comparative analysis proved that the proposed system leads to improvements in the ERV travel time when compared to the current practice of moving to the nearest edge upon detecting the ERV.

In Chapter 4, the ILP is extended to accommodate partial market penetration and larger transportation links. Optimizing the ERV intra-link movement on a large link in a single ILP optimization result in significant computation times, unrealistic for implementation in real-time systems. In addition, it may require the most downstream non-ERVs to stop at an extremely early point in time. Therefore, a sequential approach is proposed that consists of applying the ILP in a successive manner as the ERV is moving forward and as new information about downstream traffic is acquired. Each ILP takes into consideration previous ILPs’ outputs to ensure continuity of the ERV’s movement as well as consistency.
of messages that are being disseminated. Recommendations on link segment sizes and grouping of non-ERVs in the different ILP optimizations (such as setting large gaps between non-ERVs as delimiters between link segments) were developed to reduce the computation times. The sequential ILP optimization approach was tested for different market penetration levels and was compared to the current practice of moving to the nearest edge. Results showed improvements in the ERV travel time (average reduction of 3.26 seconds per 0.1 mile) as well as reductions in vehicle interactions.

In Chapter 5, the complete system that provides emergency assistance in an urban transportation network is presented. It consists of three modules: (1) an ERV route generation module to identify the best route (set of links) from origin to destination, (2) a criticality analysis module that scans the downstream non-ERVs and determines the set of downstream non-ERVs that should receive instructions messages, and (3) a sequential ILP optimization module that receives data about the critical connected non-ERVs from the previous module and that executes the ILP (previously introduced in Chapter 3 and 4) to generate the ERV maneuvering and non-ERV instruction messages. In Chapter 5, the proposed system is extended to consider the presence of intersections. In fact, the criticality analysis module adopts a heuristic approach to filter out the non-ERVs that are currently on the ERV’s remaining route but diverging before the ERV’s passage, hence avoiding the generation of unnecessary instruction messages. In addition, the ILP is extended to account for the ERV’s desired lane prior to an intersection depending on whether a right, left of through movement is to be performed. To assess the benefits and impacts of the proposed system, a sample grid transportation network with 16 signalized intersections is modeled in Netlogo, an agent-based modeling tool. Tests were conducted for different combinations of volume/capacity (v/c) ratios and market penetration levels. Improvements in ERV travel times were noted when compared with the current practice of moving to the nearest edge and for market penetration levels equal to and higher than 40%. An average of 9.09% reduction in ERV travel time is observed at 100% market penetration. Besides, this system offers tremendous benefits in terms of vehicular interactions reduction with an average of 35.46% and 81.38% for ERV/non-ERV and non-ERV/non-ERV interactions respectively at 100% market penetration. Consequently, the proposed system succeeds in providing a clearer and safer passage for the ERV as well as in improving the safety of downstream traffic.

Future works include the extension of this proposed system to accommodate multiple emergencies at the same time in the network, as this is of great need especially in cities. Besides, information about drivers’ routes (if available) will be considered while generating the instruction messages to lessen the impact of the ERV’s passage on non-ERVs movement. Recovery techniques to assist drivers while resuming their movement after the ERV’s passage will be developed. As a step forward towards deployment, assessing the implications of imperfect V2I communications is also part of future works.
Reference


