Performance Measures of Interrupted-Flow Roadways Using Re-Identification and Signal Controller Data

Christopher M. Day, Darcy M. Bullock
Performance Measures of Interrupted-Flow Roadways Using Re-Identification and Signal Controller Data

Christopher M. Day
Purdue University

Darcy M. Bullock
Purdue University

SBIR Phase 3 Joint Transportation Research Project
Traffax, Inc.
Purdue University

May 4, 2015

<table>
<thead>
<tr>
<th>Deliverable Reference:</th>
<th>D1.1 Arterial Performance Measures Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Name:</td>
<td>Sensor Fusion and MOE Development for Off-Line Traffic Analysis of Real Time Data</td>
</tr>
<tr>
<td>Contractor:</td>
<td>Traffax, Inc.</td>
</tr>
<tr>
<td>Contract Number:</td>
<td>DTFH61-14-C-00035</td>
</tr>
<tr>
<td>Contract Term Start</td>
<td>9/4/2014</td>
</tr>
<tr>
<td>Contract Term End</td>
<td>9/4/2017</td>
</tr>
<tr>
<td>Key Personnel</td>
<td>Stan Young, Darcy Bullock, Dennis So Ting Fong</td>
</tr>
</tbody>
</table>
Recommended Citation

Acknowledgments
This work was supported by Traffax/USDOT SBIR DTFH6114C00035. The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein, and do not necessarily reflect the official views or policies of the sponsoring organizations. These contents do not constitute a standard, specification, or regulation.
Table of Contents

List of Figures .............................................................................................................................................................. 3
List of Tables ................................................................................................................................................................ 7
Acknowledgments ........................................................................................................................................................ 8
Introduction ................................................................................................................................................................. 9

Module 1. Arterial Management Concepts .................................................................................................................. 11
1.1 Introduction .......................................................................................................................................................... 11
1.2 Purpose of this Report ....................................................................................................................................... 14
1.3 Organization ....................................................................................................................................................... 15
1.4 A Framework for Understanding and Controlling the System ......................................................................... 16
1.5 Defining the Arterial System .............................................................................................................................. 17
1.6 Defining Performance Objectives and Connecting Performance Measures ....................................................... 19
1.7 Network, Corridor, and Intersection Levels .................................................................................................... 23
1.7.1 50,000-ft Level: The Network ..................................................................................................................... 25
1.7.2 5,000-ft Level: the Corridor .......................................................................................................................... 30
1.7.3 500-ft Level: the Intersection ....................................................................................................................... 39
1.8 What Can Be Measured at Each Level? ............................................................................................................. 45
1.9 Implementation and Integration of Performance Measures ............................................................................. 50
1.10 Summary ........................................................................................................................................................... 52
1.11 References ......................................................................................................................................................... 53

Module 2. Introduction to New Technologies .............................................................................................................. 54
2.1 Introduction ......................................................................................................................................................... 54
2.2 Vehicle Re-Identification Data .......................................................................................................................... 54
2.2.1 Vehicle Re-Identification and Travel Time Measurement ............................................................................. 54
2.2.2 Technologies for Automated Vehicle Re-Identification .............................................................................. 56
2.2.3 Examples of Vehicle Re-Identification Travel Time Data ........................................................................... 57
2.2.4 Applications of Vehicle Re-Identification Travel Time Data ...................................................................... 62
2.2.5 Comparison to Crowdsourced Probe Data .................................................................................................. 64
2.3 High Resolution Traffic Event Data .................................................................................................................. 65
2.3.1 Traffic Control Development Process ........................................................................................................ 65
2.3.2 High-Resolution Data Concept .................................................................................................................. 66
2.3.3 Example High-Resolution Traffic Event Data ............................................................................................. 66
2.3.4 Applications of High-Resolution Traffic Event Data ................................................................................... 71
2.4 Summary ............................................................................................................................................................ 76
2.5 References ........................................................................................................................................................ 76
Module 3.  Travel Time and Travel Time Reliability .............................................................................................................. 78
  3.1  Introduction ........................................................................................................................................................................... 78
  3.2  Intersection Level Analysis ........................................................................................................................................................................... 79
    3.2.1  Traffic Flow through an Intersection ........................................................................................................................................ 79
    3.2.2  Delays of Waiting Vehicles and Pedestrians .................................................................................................................................. 88
  3.3  Corridor Level Analysis ......................................................................................................................................................................... 91
    3.3.1  Relationship between Arrival Characteristics and Travel Times .................................................................................................. 91
    3.3.2  Travel Time Sensitivity to Sensor Location .................................................................................................................................. 99
  3.4  Network Level Analysis ........................................................................................................................................................................... 104
    3.4.1  Aggregations of Lower Level Data ........................................................................................................................................ 104
    3.4.2  Analysis of Data at the Network Level ...................................................................................................................................... 109
  3.5  Guidance for Operational Analysis ............................................................................................................................................................... 113
    3.5.1  Case Study: Signal Timing Optimization over the Weekend ............................................................................................................. 113
    3.5.2  Longitudinal Tracking to Investigate Timing Plan Aging ........................................................................................................... 122
    3.5.3  Quantifying User Benefits .................................................................................................................................................... 128
  3.6  Summary ................................................................................................................................................................................................. 130
  3.7  References ............................................................................................................................................................................................... 131

Module 4.  Intersection Capacity and Demand ................................................................................................................................. 132
  4.1  Introduction ............................................................................................................................................................................................... 132
  4.2  Intersection Level ..................................................................................................................................................................................... 132
    4.2.1  Evidence of Oversaturation and Undersaturation in Microscopic Data .......................................................................................... 134
    4.2.2  Quantifying Capacity ............................................................................................................................................................. 134
    4.2.3  Quantifying Demand .......................................................................................................................................................... 140
    4.2.4  Capacity Utilization across the Intersection .................................................................................................................................. 147
    4.2.5  Utilization of Intersection Capacity under Projected Future Volumes .......................................................................................... 151
  4.3  Corridor Level .......................................................................................................................................................................................... 154
    4.3.1  A Look at Corridor Travel Times .................................................................................................................................................... 154
    4.3.2  Corridor Analysis of Intersection Utilization .................................................................................................................................. 157
  4.4  Network Level .......................................................................................................................................................................................... 161
    4.4.1  Ranking Split Failure Severity Across the Network ..................................................................................................................... 161
    4.4.2  Ranking Intersections by Capacity Utilization .................................................................................................................................. 164
  4.5  Guidance for Operational Analysis ............................................................................................................................................................... 169
    4.5.1  Measuring the Impact of a Split Adjustment .................................................................................................................................. 169
    4.5.2  Measuring the Impact of an Intersection Control Policy Change .................................................................................................. 176
    4.5.3  Investigating the Impact of Cycle Length Changes ..................................................................................................................... 181
  4.6  Summary ................................................................................................................................................................................................. 185
  4.7  References ............................................................................................................................................................................................... 186
List of Figures

Figure 1.1. Evolution of pavement management systems and arterial management systems..........................13
Figure 1.2. Basic elements of statistical process control. .............................................................................16
Figure 1.3. Functional classes of different roadway types ...........................................................................17
Figure 1.4. Breakdown of arterial management into components ....................................................................18
Figure 1.5. Analysis levels .............................................................................................................................24
Figure 1.6. Network perspective: Aggregated arterial performance .............................................................27
Figure 1.7. Network perspective: Alternative routes .....................................................................................28
Figure 1.8. Network perspective: inter-agency relationships .........................................................................29
Figure 1.9. Basic traffic signal control concepts on a corridor .................................................................31
Figure 1.10. Traffic flow in a coordinated signalized corridor .....................................................................33
Figure 1.11. Impact of poor signal timing at Int. 3 (bad offset) ....................................................................34
Figure 1.12. Importance of controller synchronization and consistent programming ...............................35
Figure 1.13. Impact of mismatched cycle length .........................................................................................36
Figure 1.14. Timeline showing the schedules of two signal programs for two adjacent corridors ............37
Figure 1.15. Impact of queue spillback ..........................................................................................................38
Figure 1.16. Twelve movements at a typical four-leg intersection ..............................................................40
Figure 1.17. Theoretical framework for bulk service of queued demand with steady arrivals ..................41
Figure 1.18. Impact of fluctuations in green times and arrival rates ..........................................................42
Figure 1.19. Illustration of a split failure ........................................................................................................43
Figure 1.20. Illustration of starvation ..............................................................................................................44
Figure 1.21. Network-level performance analysis: Comparison of alternative routes [13] .......................47
Figure 1.22. Corridor-level performance analysis: Day-to-day travel times along a corridor [2] .............48
Figure 1.23. Intersection-level performance analysis: Split failures by movement ....................................49
Figure 2.1. Travel times from vehicle re-identification ..............................................................................55
Figure 2.2. Example data from an automated vehicle re-identification travel time study [3,8] .................57
Figure 2.3. Test corridor for vehicle re-identification travel times ..............................................................59
Figure 2.4. Raw travel time data for northbound vehicles from a 6-month time period ...............................60
Figure 2.5. 24 hours of raw data for northbound vehicles on Wednesday, 10/26/2011 ............................60
Figure 2.6. Distribution of travel times for northbound vehicles during the AM peak (6:00-9:30) ..........60
Figure 2.7. Identifying outliers in arterial travel time data ...........................................................................61
Figure 2.8. Travel time cumulative frequency distributions before and after optimization of signal timing [9] ..........63
Figure 2.9. Comparison of performance and proportion of observed vehicles on alternative signal timing [10] ..........................63
Figure 2.10. The traffic control development process [13] ..........................................................................65
Figure 2.11. 24-hour plots of effective cycle length .....................................................................................70
Figure 2.12. Volume-to-capacity ratio calculation concept ............................................................................72
Figure 4.3. A comparison of estimated capacity to observed capacity for ring 1 (phases 1–4) for one cycle [2]......136
Figure 4.4. Measuring capacity. .................................................................138
Figure 4.5. Green time and capacity of a single phase. .................................139
Figure 4.6. Capacities by phase assuming a 1900 veh/h/lane saturation flow rate.141
Figure 4.7. Volumes by phase. ...................................................................141
Figure 4.8. Volume-to-Capacity ratio by phase. ...........................................143
Figure 4.9. CFD of v/c ratios over 24 hours by phase. ..................................143
Figure 4.10. GOR and ROR calculation for a single cycle of an oversaturated movement [3]..........................145
Figure 4.11. Plot of Red Occupancy Ratio (ROR) versus Green Occupancy Ratio (GOR) ......................................146
Figure 4.12. ROR/GOR by phase. ...............................................................148
Figure 4.13. Number of split failures per half hour, by phase. .....................148
Figure 4.14. Degree of intersection saturation............................................150
Figure 4.15. Number of split failures per half hour, by phase, sorted by the degree of intersection saturation (using a threshold value of 0.75) into “correctable” and “not correctable” categories .................................................................150
Figure 4.16. Change in the degree of intersection saturation under different volume growth rates over 10 years...152
Figure 4.17. Spare capacity by time of day under different volume growth rates over 10 years..............................152
Figure 4.18. Change in the degree of intersection saturation over time under 2% annual growth. .......................153
Figure 4.19. Spare capacity by time of day over time at 2% annual growth. ..................................................................................................................153
Figure 4.20. Dimensions of analysis for a corridor [4]..................................154
Figure 4.21. Map of a four-intersection span of SR 37 in Noblesville, Indiana. ..............................................................155
Figure 4.22. Comparison of weekday arterial travel times from different time of day periods. .................................156
Figure 4.23. Degree of intersection saturation by intersection (10-cycle moving average).........................................158
Figure 4.24. Rank-ordered degree of intersection saturation by intersection. (Data originally presented in [1]). .....158
Figure 4.25. Split failures per hour by phase across the corridor. (Data originally presented in [1]). .........................160
Figure 4.26. Agency-wide analysis: Maximum phase failure rate by intersection. Data is shown for 61 intersections with data from May 31, 2014 [5]........................................................................................................163
Figure 4.27. Agency-wide analysis: Degree of intersection saturation (Xc) for the PM peak hour (1700-1800). .....165
Figure 4.28. Agency-wide analysis: Degree of intersection saturation (Xc) for the PM peak hour (1700-1800). Close-up of the Indianapolis area. .................................................................166
Figure 4.29. Agency-wide analysis: Pareto sorted list of intersections by degree of intersection saturation (based on afternoon peak hour volumes), under different projected annual growth rates after 10 years. ......................168
Figure 4.30. Agency-wide analysis: Pareto sorted list of intersections by degree of intersection saturation (based on afternoon peak hour volumes), under different number of years at 2% annual growth. .................................................................168
Figure 4.31. US 31 and 126th St. [3]...............................................................170
Figure 4.32. Capacity utilization performance measures for an example intersection [3]........................................171
Figure 4.33. Split adjustments at US 31 and 126th St [3]................................172
Figure 4.34. Number of split failures before and after split adjustment [3] .................................................................173
Figure 4.35. Volume-to-capacity ratios before and after a split adjustment [3]. ........................................................................173
Figure 4.36. PCDs for time periods before and after the offset adjustment, for both directions. ..........................175
Figure 4.37. Explanation of the operational change. .................................177
Figure 4.38. Graphs of cumulative frequency of V/C ratios. Arrows indicate change in the average value [7]......179
Figure 4.39. Impact of early yield on progression [8]. ..................................180
Figure 4.40. A five-section span of SR 37 [10]. .........................................182

Arterial Performance Measures Report
Contract #: DTFH61-14-C-00035
Figure 4.41. Percentage of phases forced off at different cycle lengths [10]............................................................ 183
Figure 4.42. Travel time distributions[10]................................................................................................................. 183
List of Tables

Table 1.1. Example of possible objectives and how they might be prioritized between two different stakeholders. ................................................................................................................................................................ 22
Table 1.2. An example of matching potential performance measures against possible objectives.................. 22
Table 1.3. Personnel roles in arterial management and opportunities for performance measures to assist and improve tasks, and new roles (highlighted) required for successful deployment of performance measures. ...........51
Table 2.1. Example of high resolution signal data ...................................................................................................... 68
Table 3.1. Annualized savings due to signal retiming [11]........................................................................................ 129
Table 4.1. Example of agency-wide split failure data: Top 30 intersections, sorted by maximum percentage of failures during the PM peak (1500-1900), May 31, 2014 [5]. ................................................................. 162
Acknowledgments

This work was supported by Traffax/USDOT SBIR DTFH61-14-C-00035. The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein, and do not necessarily reflect the official views or policies of the sponsoring organizations. These contents do not constitute a standard, specification, or regulation.
Introduction

Objectives

Currently, transportation agencies are increasingly asked to demonstrate accountability for the systems that they manage through outcome-based performance measures. The most recent comprehensive highway funding bill, MAP-21, contained several provisions that have driven this development, as well as current funding conditions for transportation agencies that have necessitated better targeting of resources. Performance measurement is essential for validating whether those resources are being invested as close to optimally as possible.

In highway operations, measuring the performance of signalized arterials has historically been quite challenging because of the dynamic nature of the traffic control. In the past, the analysis consisted of feeding macroscopic data (such as 15-minute vehicle volumes) into analytical models that, if calibrated correctly, would provide an estimate of the system performance. Recently, however, some innovations have arisen that now enable that performance to be measured rather than modeled. These include travel time measurements through vehicle re-identification, and traffic event data that records, in real time, detailed timestamps of vehicle detections, signal state changes, and other events that occur at an intersection. This report presents performance measures based on these data sets to measure success at the achievement of various objectives in arterial performance.

Approach and Organization

The report approaches the subject of arterial performance from a series of perspectives that vary according to the “level” at which the analysis is taking place, varying from microscopic to macroscopic. The three levels defined in the report are the intersection, corridor, and network (“500 ft”, “5000 ft”, and “50,000 ft”). Highly detailed information about intersection operations can be developed from analysis of the traffic event data, revealing intricacies about the performance of individual movements at intersections. This data can be aggregated to evaluate the overall performance of corridors and networks. The travel time data, in the meanwhile, can directly measure the performance at each level based on defining the endpoints and waypoints in the vehicle re-identification process.

The report is organized into four modules:

Module 1. Arterial Management Concepts

The first step in engaging in a performance measurement activity is to define the objectives that are to be addressed. These vary according to the priorities of the agency and the context of an arterial system. This module focuses on objectives by presenting fundamental operating concepts for arterial systems at the network, corridor, and intersection levels, and exploring what can be measured at each level.
Module 2. Introduction to New Technologies
This module introduces readers to the concepts behind vehicle re-identification and high-resolution event data. The information captured by these technologies are explained, and some example applications provide a preview of potential use cases for the data sets.

Module 3. Travel Time and Travel Time Reliability
A key aspect transportation system performance is the amount of time required for users to make it through the system from an origin to a destination. This module focuses on measuring and evaluating travel times within an arterial context. Concepts related to the performance at intersections, through corridors, and across networks are explained. A series of case studies are presented in which the data is used to measure the impact of changes to the signal timing, and quantifying the resulting user benefits.

Module 4. Intersection Capacity and Demand
The interaction between capacity and demand at intersections is important to many objectives in arterial operations. The efficient allocation of capacity at intersections requires balancing the distribution of green times according to demands. Beyond a certain point, however, the overall level of demand exceeds the maximum potential capacity of the intersection, and capital improvements are needed. This module presents performance measures that measure the amounts of capacity supplied and the degree to which it is utilized. Combining these concepts enables differentiation between operational deficiencies that can be corrected through adjustment of signal timing, and those requiring larger investments. Case studies are presented where the performance measures supported investigation of different operational strategies.
Module 1. Arterial Management Concepts

1.1 Introduction

Most transportation agencies operate and maintain an inventory of roads for a particular jurisdiction, and as public sector organizations, they are asked to demonstrate accountability for the performance of those roads. In the past, congestion management programs (CMPs) required by statute, have forced agencies to develop some existing methods of measuring performance, which have relied on manual data collection methods. Presently, agency resources are decreasing while the demand for better infrastructure is increasing. As a consequence of this situation, agencies are facing increased pressure to show that the resources that have been provided are being used efficiently. This was clearly demonstrated in the most recent broad transportation funding initiative, MAP-21, which contained several provisions that aimed to increase the measurement of performance in transportation [1].

At the same time, the proliferation of connected computing devices such as smart phones has begun to influence the perceptions and expectations of the traveling public. There is a demand for more and better information, particularly in the transportation industry, whose performance directly affects the travel behavior of individuals and the activities in which they participate—in other words, the overall economic activity in a location.

There are several different aspects of transportation system performance, including infrastructure quality, safety, mobility, accessibility, environmental impact, and cost-effectiveness. All of these aspects are important in their own right. However, mobility is more immediately perceived by the users of the system—the customers of the agency. The relative ability of a transportation system to move individuals and goods from one location to another is the strongest economic impact attributable to a transportation system. The day-to-day operations of the system profoundly affect its mobility performance. However, because of the difficulty in measuring mobility, methods for quantifying that performance are not as developed as some other aspects.

For example, in roadway systems, there are mature, data-driven methodologies in widespread use for measuring pavement quality. Many agencies have a pavement management system that they have been using for many years. The basic research needed to bring about pavement management systems required several decades to develop, as illustrated by Figure 1.1. Basic exploratory research on the performance measures to be used, and development of consensus on those performance measures, were first required. After this, standard methodologies for data collection came into place, followed by the development of systems for handling the data. The final stages involved the integration of those systems into engineering practices, and finally decision making on investment of resources. While there remain many factors that contribute to the decision on whether to repave and reconstruct roadways, pavement management systems play an important role in informing the decision making process for engineers and administrators.
Arterial management systems are not as well developed, partly because the technologies to develop the data sets needed to facilitate creation of systems have only recently become available on a widespread basis. Past practices were mostly limited to manual and software-assisted methodologies based on analytical models that would estimate roadway performance using a series of input volumes and details about roadway configuration and intersection control. These estimates are limited by the fidelity of the models and the accuracy of the input volumes and the accuracy by which the analyst has characterized the roadway by calibration of the model parameters. Turning movements in particular have historically been labor-intensive to obtain. While automatic data collection has been used on freeway facilities, it has not been possible to directly translate these methods to non-freeway facilities because interrupted flow is inherently more complex.

In the past several years, innovations in traffic technology have enabled a more rigorous analysis of traffic. The two new data sources focused on in this report are vehicle re-identification travel times and high-resolution traffic event data. These new technologies enable the engineer to monitor and visualize arterial operations at adjustable levels of detail, depending on the use case. Trends in performance can be observed over time and across a wide area, or individual events may be examined when the details are needed. Previously published research has demonstrated the utility of combining the two data sets to optimize arterial operations [2, 3].

Momentum has been growing in the industry to rapidly adopt these technologies, as demonstrated by a growing number of vendors now offering data collection capabilities in traffic control systems, and a growing national emphasis on performance monitoring in the transportation industry, especially with MAP-21. The USDOT has articulated several provisions for facilitating of performance measures for the National Highway System, including the establishment of “data elements… necessary to collect and maintain standardized data to carry out a performance-based approach”. [1] Performance monitoring has similarly been incentivized in planning activities. A proposed FHWA rule creates a mandate for MPOs and state DOTs to begin using performance-based planning and programming practices, to set performance targets, and to establish performance measures that quantify whether those targets have been successfully attained. All of these developments point to the need to accelerate the implementation of arterial performance measurement in practice.
Figure 1.1. Evolution of pavement management systems and arterial management systems.
1.2 Purpose of this Report

This report presents a framework for data-driven, performance-based management of arterial highways. The report presents a series of technical definitions, guides for interpretation, and methods of implementation for a portfolio of performance measurements that use vehicle re-identification and high-resolution traffic signal event data. The report guides practitioners through the steps of:

- Selection of performance measures, by connecting those metrics to agency objectives;
- Implementation of performance measures, by providing basic technical information on the performance measures; and
- Usage of performance measures for decision-making in operations and maintenance as well as planning.

The performance measures identified in the report span the breadth of arterial management from a single intersection, to a corridor comprising multiple intersections, and lastly to a network comprising multiple corridors. The objectives of the agency are expressed differently with respect to the system elements in play at each the intersection, corridor, and network level. Thus, central concepts include connecting performance measures to agency objectives and differentiating how these are implemented according to the scale of the analysis.

This report focuses primarily on the quality of the mobility provided in signalized arterials, how to measure aspects of mobility, and how to make use of this information to improve system operations. A methodology is developed which uses measurements of actual traffic performance to document conditions on an arterial at the intersection, corridor, and network levels. Finally, the report focuses on how this information can be used to improve arterial operations and to inform the planning process.
1.3 Organization

This report is organized into four modules that can be independently used to facilitate better understanding of each topic. While some concepts are explained in more detail in certain modules, each module can be presented independently as a source of information for training engineers and planners in using the methodology or briefing administrators and decision-makers.

Module 1: Arterial Management Concepts. The first module focuses on articulating agency objectives and connecting these to performance measures. These concepts represent a foundational basis for the rest of the framework. Because different metrics are needed to characterize performance at different levels, the presentation of concepts and performance measures in other modules is structured by a series of incrementally detailed perspectives. These are identified as the Network, Corridor, and Intersection levels.

Module 2: Introduction to New Technologies. The data sets used to develop a series of performance measures in this report are relatively new. This module will assist readers who are unfamiliar with these data sets by presenting an introductory discussion. The types of information made available by vehicle re-identification and high-resolution traffic event data are presented. The module describes how system monitoring can be considerably enhanced on an unprecedentedly broad scale using these data sources.

Module 3: Travel Time Reliability. The amount of time needed to get from one point to another in the network relates to many different transportation system objectives. This module approaches the performance of an arterial from the perspective of the user, by assessing performance based on observed travel times and their reliability. This is accomplished by sampling travel times based on vehicle re-identification at different locations in the network, and by understanding operational details related to traffic flows through intersections as monitored by traffic event data. A series of performance measures are developed for network, corridor, and intersection levels using vehicle re-identification and traffic event data. The module includes example case studies demonstrating the use of the performance measures.

Module 4: System Capacity and Demand. The relationship between supply of capacity and the traffic demand applies to numerous objectives. In an arterial system, capacity is typically most constrained at intersections. Effective capacity allocation is achieved by a balanced distribution of green times at traffic signals, and by sufficient provision of turning lanes and other geometric facilities. This module presents a series of performance measures based on traffic event data and vehicle re-identification data to explore capacity utilization at the network, corridor, and intersection levels. These allow for differentiation between more readily correctable deficiencies of signal timing, and those which would require addition of capacity or changes to vehicle routing. The module includes example case studies demonstrating the use of the performance measures.
1.4 A Framework for Understanding and Controlling the System

In the manufacturing industry, small inefficiencies can be extremely costly to the overall operation. The strategy of statistical process control (SPC) has been developed to reduce these inefficiencies and minimize the possibility of failure. Transportation systems, and traffic control in particular, is far less predictable than a factory environment, but the central concepts of SPC can still serve as a framework for developing a data-driven system for arterial management.

The five key strategies of SPC are illustrated in Figure 1.2: Define, Measure, Analyze, Improve, and Control (DMAIC). These form a continuous, cyclic process. These concepts can be applied to arterial management as follows [4]:

- **Define**. Articulate and understand the agency’s objectives and identify the target levels of operation that are desirable to meet or exceed.
- **Measure**. Obtain the field data needed to completely characterize the operation of the system.
- **Analyze**. Distill the data into useful information (performance measures) that directly respond to the defined objectives.
- **Improve**. Use the results of the analysis to make the system work better by identifying specific aspects of the operation that need to be changed.
- **Control**. Proceed with the improvements and use these to better inform objective definitions for the future.

![Figure 1.2. Basic elements of statistical process control.](image-url)
1.5 Defining the Arterial System

The DMAIC concept shown in the last section begins with the step of defining, which is primarily the focus of this Module—with measuring, analyzing, improving, and controlling the subjects of subsequent sections. To begin defining, it is helpful to consider the context of arterial systems and the components that make up an arterial.

As a functional class of roadways, arterials often serve a dual purpose of providing both mobility and accessibility, as illustrated by Figure 1.3. The dominant function of an arterial varies according to factors such as the neighboring land use, presence of multimodal facilities, and density of intersections. In some environments, an arterial serves as a high-speed road that functions more or less as a freeway with at-grade intersections, thus prioritizing mobility. Elsewhere, arterials carry a heavy amount of traffic at slower speeds yet must also adequately serve crossing movements, pedestrians, transit, and other users. In many agencies, arterials often have heterogeneous characteristics along their length, with variations occurring as the roadway passes through urban, suburban, and rural areas.

The primary distinguishing characteristic of an arterial system is the presence of at-grade intersections where traffic flow is interrupted. At intersections, traffic control devices manage the right-of-way. On most arterials, the “mainline” movements carrying traffic on the primary directions along the route are typically prioritized. Most intersections function under two-way stop control and traffic signals. Roundabouts and all-way stop control are also sometimes used on arterials where accessibility is the dominant function.

![Figure 1.3. Functional classes of different roadway types.](image-url)
Figure 1.4 examines the overall problem of arterial management and considers the various components of the system. Each element in the figure depends on the functionality of the elements to the right. The top level management functions, for example, depend on the operational functions, which themselves depend on having working components. The operations level divides the purpose of an arterial into two categories:

- The *system* category of operations involves the functioning of groups of traffic signals as an organized structure. This requires the availability of working communication, and of optimally configured system timing plans.
- The *local* category of operations involves the switching of service states—rotation through phases—at each individual intersection. This requires the availability of working detection, and of optimally configured intersection control plans.

These two categories respectively represent the two competing tasks of arterial operation, signal progression and capacity allocation.

- *Progression* refers to the ability of a traffic signal system to move along routes in the system efficiently. That is, traffic is progressed through successive intersections without stopping. This is a vital function of an arterial roadway.
- *Capacity allocation* refers to the qualities of efficiency and equitability in the handling of competing demands for service at an intersection—among conflicting movements serving alternative modes of traffic.

Most of the mobility oriented objectives of arterial operation will be related to one or both of these primary categories.
1.6 Defining Performance Objectives and Connecting Performance Measures

The first stage in defining an arterial management system is to articulate the objectives of the agency’s arterial program. These objectives will vary in their priority among the different stakeholders involved. While objectives may evolve over time as the arterial network and the communities that it serves both change, having some idea of the focus areas to be emphasized will steer the performance measures to a purposeful end.

From an agency perspective, arterial performance objectives can be roughly divided into two broad categories: user-oriented and agency-oriented objectives.

Roadway users include motorists and passengers, commercial vehicles, bicyclists, and pedestrians. The users are directly affected by investments in and strategies used to operate the system. The level of satisfaction of the users may be stated as the quality of service provided by the arterial network. Ideally, users traveling from an origin to a destination in the network experience fewer delays and other inefficiencies. The articulation of these objectives varies according to the type of user, which is defined by mode and route [5]. Example user-oriented objectives include:

- Maintain smooth traffic flow by scheduling and sequencing service (green) times at neighboring signalized intersections.
- Minimize the waiting times of stopped traffic. Distribute capacity as fairly and efficiently as possible across movements and modes.
- Provide adequate service times for non-motorized modes at signalized intersections.
- Provide the ability for transit vehicles to execute loading and unloading activities and prioritize transit vehicle service to expedite travel times.
- Minimize the amounts of pollution and noise by improving the efficiency of vehicle movement through the system.

The second category concerns the agency that owns and operates the facility. The agency has a need to ensure that its resources are expended in the best possible way. This means that inefficiencies should be identified and that strategies should be evaluated. In addition, performance measures are important for making informed decisions about long term capital investments. Decisions regarding whether or not to make an investment or resource allocation directly relate to agency resource levels. In essence, the agency must be able to sustain operations at desired performance level over the design life of a facility. New investments in the system require resources to construct, operate, and maintain. System operators are interested in understanding system performance to make better informed planning decisions, as well as to quantify the return on past investments. Example agency-oriented objectives include:

- Determine attainment of performance targets.
- Ensure that investments provide a sufficient return.
- Ensure that personnel and other resources are used optimally.
- Develop data sets for planning, incident management, and public relations.
- Influence the traveling and/or route choice behavior of the public.
User-oriented and agency-oriented concerns converge in the operation of the system, in the planning process, during selection and evaluation of projects, and in public discussion of the state of transportation. By pairing the appropriate performance measure with an articulated objective, these conversations can evolve from a subjective expression of how the system seems to be doing, to an objective evaluation of how the system actually is doing. The following scenarios contemplate how this might actually come to pass for an agency.

- An agency wants to set reasonable performance targets for its inventory of arterial corridors to evaluate the overall system performance. Performance measures related to travel time reliability and progression define thresholds of well-performing and poorly performing arterial corridors. This classification scheme allows the agency to set the long-term performance goals of maintaining “good” performance on 80% or greater of corridors, while not allowing more than 5% to fall into the “poor” performance category. At the same time, the agency is tasked with promoting “livable” communities, and some arterials pass through such areas. An analysis of traffic event data enables the agency to dually evaluate travel time objectives along with the quality of performance for pedestrians and bicyclists.

- Given limited operations funding, an agency wants to make better use of what resources it has to manage its arterial network. Using vehicle travel time distributions and arrival on green characteristics, the agency’s inventory of corridors is sorted and ranked by performance so that those corridors having the most actionable deficiencies are prioritized for operational improvements. Subsequently, corridors are ranked to efficiently program the limited budget. By implementing data collection, the established policy of retiming corridors on a five-year schedule can be altered to a performance-based policy, so those corridors which need retiming less frequently can be set aside for future action and those needing more immediate action can be expedited.

- The planning department is assessing the impact of development proposals anticipated to change traffic patterns of a city center. The anticipated increase in demand can be assessed against existing (and projected) arterial capacity reserves to determine if the current arterial network can accommodate the increase in demand with operational improvements, or whether capital improvements will be needed.

- The agency has invested in a major capital improvement investment to decrease congestion on a major corridor. The administration wants the transportation department to measure the impact of the improvements in a before/after scenario. Travel time and travel time reliability performance measures are applied before and after the capital improvement to reveal the improvement in mobility, while traffic event data is used to characterize whether the changes have been beneficial or detrimental to the performance of side-street and crossing movements.

The preceding examples represent a handful of the potential types of objectives that could be articulated. These will have different expressions and levels of priority according to the agency and its relationship to the public that it serves, and their priorities, as well as the context of the arterial.
Table 1.1 presents an example list of objectives that could be identified through discussion with stakeholders in the operation of an arterial system, and the relative priorities of the two primary stakeholders. In this example, Stakeholder “A” may represent a state transportation agency, whose primary concerns are providing high quality progression of traffic into and out of major cities. Stakeholder “B” may represent an MPO or local transportation agency for whom progression is important, but not as high of a priority as improving pedestrian service and providing equitable amounts of green time. Alternately, the two stakeholders could represent traffic engineers and planners within the same agency. It is important to understand the priorities of each stakeholder when developing an arterial management system so that multiple objectives can be met.

After objectives have been identified, they need to be matched with performance measures. This process creates a plan for the types of performance measures to be obtained with the arterial management system, and identifies what can feasibly be measured. Table 1.2 shows an example using the same list of objectives presented earlier. Here, they are matched against a candidate list of performance measures that could potentially be collected in an automated fashion using vehicle re-identification and high-resolution data. The remaining portions of the report explain the meaning of these in more detail.

These two tables are examples, first and foremost—many additional objectives could be included in the list, and other performance measures could also be added to the right of the existing columns. However, the process represented in this example is the fundamental step for establishing the central framework of an arterial management system. The matrix that matches the core objectives of the operation against measurable elements is the kernel around which the rest of the system needs to be developed. For most objectives pertaining to arterial operations and planning, the success of the agency at attaining the desired performance can be quantified through measurement. The data sets described in this report allow a much more detailed analysis than could otherwise be obtained.
Table 1.1. Example of possible objectives and how they might be prioritized between two different stakeholders.

<table>
<thead>
<tr>
<th>Objective</th>
<th>Stakeholder A</th>
<th>Stakeholder B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improve Progression</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Improve Capacity Allocation</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Improve Pedestrian Service</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Improve Bicycle Service</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Maintain Working Detection</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Maintain Working Preemption</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Minimize Pollution and Noise</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Automate Traffic Counts</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Develop Origin-Destination Data</td>
<td>9</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 1.2. An example of matching potential performance measures against possible objectives.

<table>
<thead>
<tr>
<th>Objective</th>
<th>Traffic Volumes</th>
<th>Green Time</th>
<th>V/C Ratio</th>
<th>Occupancy Ratio</th>
<th>Percent on Green</th>
<th>Coordination Diagram</th>
<th>Travel Time</th>
<th>Detector Error Alarms</th>
<th>Preemption Durations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improve Progression</td>
<td></td>
<td>X X X X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improve Capacity Allocation</td>
<td>X X X X</td>
<td>X X X X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improve Pedestrian Service</td>
<td>X X X X</td>
<td>X X X X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automate Traffic Counts</td>
<td>X X X X</td>
<td>X X X X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimize Pollution and Noise</td>
<td>X X X X</td>
<td>X X X X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop Origin-Destination Data</td>
<td>X X X X</td>
<td>X X X X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1.7 Network, Corridor, and Intersection Levels

This report focuses its attention on objectives related to mobility in a signalized arterial system. Mobility relates to several dimensions of performance that vary by mode and the spatial perspective of the analysis. To convey this concept, this report considers three different analysis “levels” that go from a more holistic to more detailed views of performance. These are illustrated by Figure 1.5.

- The network or “50,000 ft” level (Figure 1.5a) encompasses the entire arterial program of an agency. At this level, concerns are focused on budget layout for operations and maintenance and for making capital improvements throughout the arterial network. Interactions with freeway networks and with other agencies are also relevant at this level. Measurable aspects of performance include aggregated performance of smaller components, and the quality of service on selected routes through the system. The travel time from a particular origin to a particular destination would be an example.

  The network in Figure 1.5a shows a web of streets in a medium-sized city that interconnect with each other and with a freeway. For a state agency, a district may encompass numerous arterials that are not necessarily interconnected, but which interact with local agencies and likely with the freeway system.

- The corridor or “5000 ft” level (Figure 1.5b) is compatible with a particular arterial corridor or a route with multiple intersections. Agency concerns include travel time characteristics along the corridor, the existence of critical capacity deficiencies, and maintenance of consistent operation. These are all measurable aspects of performance at the corridor level. In some cases, parallel and crossing routes may be a concern. In Figure 1.5b, the example corridor includes a few crossing arterials with intersections relatively close by that would need to be coordinated to ensure efficient operation. The corridor also includes an interchange with a freeway route.

- The intersection or “500 ft” level (Figure 1.5c) is where the performance of individual movements becomes relevant, especially the movements that impact the overall intersection and cascade upward to affect corridor or network performance. Occurrences of demand-capacity mismatches, such as spillbacks caused by inadequate capacity or excessive demand, are measurable at the intersection level, as will be demonstrated in this report. As Figure 1.5c shows, the definition of “intersection” might sometimes be extended to include multiple road crossings that function as a single control unit (such as a single controller diamond interchange).
Figure 1.5. Analysis levels.
1.7.1 50,000-ft Level: The Network

The network level perspective, as the system might be viewed from an altitude of 50,000 feet, comprises the entire arterial program of a transportation agency, or perhaps of a district belonging to a large, geographically dispersed agency. The overall performance of the network results from the collective performance of the many individual system components. At a system level, the primary concern is to identify those components out of the overall inventory that are performing poorly, and to develop a complete picture of the quality of service through the network, for deciding where to invest resources and to program the agency budget for arterial highways.

In contrast to freeways, arterial performance is determined chiefly by intersection performance, because intersections constrain capacity. Whereas each lane of roadway with traffic moving at free flow speed has a capacity of about 1900 to 2200 vehicles per hour per lane, this can only be achieved for a certain portion of the time on an arterial because of traffic control. The red indications take away a certain percentage of the capacity.

Much of the outcome-related performance analysis that can be conducted at this level would result from aggregating upward from the lower levels. Performance data from intersections and corridors are compiled to produce the overall picture. Figure 1.6 shows the example arterial network with performance displayed and ranked in a qualitative manner, showing locations of poor performance in the system. In this example, some corridors are defined and other intersections are displayed in isolation.

The actual performance measures and their display would vary in implementation, but this conceptual illustration demonstrates the purpose of these graphics on a network level. In this case, the worst performing area stands out clearly as the red patch near the confluence of two major arterials. The display might show, for example, the prevalence of excessive queuing where “acceptable” performance represents minimal queue formation and “very poor” indicates spillbacks. In this case, the red-colored region shows the location of a critical bottleneck in the overall system. This sort of qualitative display appears in numerous existing map applications; however, that data is oriented toward public information and not necessarily agency objectives.

Interaction with freeway systems is also a concern at the network level. Figure 1.7 shows a view of the example arterial network shown earlier. Here, a route along an Interstate highway is highlighted in blue, and an alternate route using the arterial network is shown in green. If an incident were to occur on the Interstate between the starting and ending interchanges, a heavy amount of traffic could be diverted to the arterial route. The added traffic would complicate operations that are set up to handle the local traffic. The diversion route differs from the usual priority route, causing substantial disruptions, especially where the diverted traffic executes turning movements.

Cross-jurisdictional interactions of traffic are also relevant at the network level. The example arterial network shown earlier is repeated in Figure 1.8. Here, the signalized intersections are color-coded to indicate the responsible agency. Four different agencies are included: the state DOT, the county DOT, and two different municipalities (separated by the river). Arterial corridors or other groups of intersections that operate in coordination are illustrated by linked...
color bands. These cross each other frequently, with some jurisdictions’ arterial routes separating others into two groups. There are numerous locations where coordinated operations between the signalized intersections would require substantial cooperation between their respective agencies.

Successful network level performance measures require a good coverage of the arterial system with data collection instrumentation. Establishing this infrastructure requires an investment of resources, although as discussed in Module 2, it can be added modularly and inexpensively, as has been done by several state and local agencies.
Figure 1.6. Network perspective: Aggregated arterial performance.
Figure 1.7. Network perspective: Alternative routes.
Figure 1.8. Network perspective: inter-agency relationships.
1.7.2  5,000-ft Level: the Corridor

The next level down from the network is the corridor, which might be described as the view from 5000 ft. A “corridor” spans multiple intersections that are physically connected along an arterial roadway, or associated with a route through the arterial network. If the intersections are signalized, they may operate as a coordinated system. At the corridor scale, provision of smooth traffic flow along high priority routes is a common objective. For example, many agencies prioritize travel into and out of areas where workplaces are concentrated; typically this means a coordination plan emphasizing inbound flow in the morning and outbound flow during the evening. Other objectives can be stated depending on the context of the corridor. For example, an arterial that carries traffic to and from a freeway interchange might be better served by a plan that prioritizes vehicles entering and exiting the freeway.

Figure 1.9 illustrates the basic concepts of coordinated signal operations using a time-space diagram. This graphic shows distance along the vertical axis, and time along the horizontal axis. The green or red state of the northbound movement at each intersection is shown by the green and red horizontal bars. The timing plan establishes a “green band” that goes through all four intersections, as illustrated by the dashed green lines. The slope of these lines corresponds to the travel speed along the corridor. The three basic signal timing concepts [5] are also illustrated:

- The **cycle length** is the amount of time needed for the intersection to rotate through all of its movements.
- The **offset** is the displacement of the cycle at a particular intersection relative to the “system cycle,” which establishes a reference point to coordinate all of the intersections. In Figure 1.9, intersections 1 and 4 have an offset of zero, while Int. 2 and Int. 3 have nonzero offsets that displace their timing from the reference point.
- For purposes of this discussion, the **split** is essentially the amount of time designated within the cycle for different groups of movements at an intersection. In Figure 1.9, the split that serves the northbound coordinated movement is half of the cycle length.

These terms have very specific operational meanings in coordinated signal control, but they represent basic concepts that apply to all forms of signal control. When the green times are actuated, their durations will vary, so it no longer becomes possible to explicitly see the cycle, offset, and split in a view such as Figure 1.9, but they still exist in the background. Further, in the absence of a coordination scheme, there will not be a fixed cycle length, but each intersection will still require a certain amount of time to serve all of its demand, and the start of green at one intersection will still have a relationship relative to that of the neighboring intersection, although it will be much less predictable.
Figure 1.9. Basic traffic signal control concepts on a corridor.
Figure 1.10 shows an example time-space diagram along one direction in a corridor with four signalized intersections. The red bars represent times when the traffic signals are red for that direction. The black lines represent the trajectories of individual vehicles, which move upward on the diagram through the four successive intersections. This particular time-space diagram illustrates a situation where the signal coordination is, for the most part, rather successful.

- Arrivals at the entry point into the system occur randomly (callout “i”), when there are no upstream traffic signals within a close distance. These random arrivals are grouped into platoons through interaction with the first signal.
- In addition to vehicles entering along the roadway, other vehicles will enter from side streets and driveways (callout “ii”).
- Queues of stopped vehicles are represented by the horizontal trajectories that occur at various locations in the diagram, usually in front of a red signal, taking on a characteristic triangular shape due to queue formation and dispersal (callout “iii”). The dashed lines show the “shockwaves” associated with the growing queue of vehicles stopping on green and the discharge of the queue after the start of green.
- Ideally, the signal timing will be configured to allow platoons of vehicles to pass through a series of signals without having to slow or stop. Callout “iv” shows the successful creation of a “green band” that successfully carries a platoon of vehicles through Int. 2, 3, and 4. The vehicles arrive on green at each of these intersections.
- Otherwise ideal progression is sometimes impeded by variations in the starting and ending times of green times, formation of large queues, and other occasional anomalies. Callout “v” shows a situation where the platoon is larger than in previous cycles and the start of red cuts it in half. The resulting queue slows down the arriving platoon in the next cycle (callout “vi”).
Figure 1.10. Traffic flow in a coordinated signalized corridor.
For the most part, signal coordination successfully progresses traffic in Figure 1.10. For contrast, Figure 1.11 illustrates a situation where the offset at Int. 3 has been displaced by about half of a cycle length, so that it no longer lines up to form a green band. Vehicles arrive randomly at Int. 1 (callout “i”), and form queues as they depart (callout “ii”). These platoons typically arrive on green at Int. 2. However, at Int. 3 they usually arrive on red (callout “iii”) during every cycle, forming queues that are cleared during the next green. Many of the vehicles departing Int. 3 are forced to stop again at the next intersection (callout “iv”). The net effect on corridor operations is an increase in travel times.

Figure 1.11. Impact of poor signal timing at Int. 3 (bad offset).
Another issue is controller synchronization. Coordination requires that neighboring intersections operate under compatible timing plans. Typically, this means that the two intersections must use the same background cycle length in order to line up the green bands during a succession of cycles [5]. Figure 1.12 illustrates the situation when a cycle length mismatch occurs. Figure 1.12a shows the ideal case, where the two neighboring intersections have the same cycle length and their greens line up to form a band during each cycle. Figure 1.12b illustrates a cycle length mismatch. Because the upstream intersection has a smaller cycle length, the green bands initiated at that intersection only occasionally line up with the greens at the downstream intersection.

(a) Consistent cycle length.

(b) Mismatched cycle length.

Figure 1.12. Importance of controller synchronization and consistent programming.
Figure 1.13 illustrates the impact of a cycle length mismatch. Here, the cycle length at Int. 3 is longer than the cycle length at the other intersections. This scheme occasionally results in good progression, more or less by accident (callout “i”). However, more often, vehicles arriving at Int. 3 are stopped (callout “ii”) and vehicles released from Int. 3 are stopped at the next intersection (callout “iii”). The net effect on the corridor performance is an increase in travel times, with much more variation (i.e., less reliability) in those travel times.

Figure 1.13. Impact of mismatched cycle length.
The cycle length mismatch is an extreme example of corridor performance under disorganized traffic control, and it occurs more commonly than might be expected. One reason may be as simple as a mistake in configuring the time-of-day plan in a controller, or setting the controller to the wrong time zone. Controller clocks tend to drift, and must be periodically re-synchronized. A loss of communications will prevent this from happening; a drift of a several seconds is equivalent to a change in the offset. In some situations, two separate corridors neighboring each other may run different timing plans. This can happen when two independent corridors grow together over time by addition of new traffic signals on either end. Figure 1.14 shows an example of two neighboring systems running quite different time-of-day plans. It would be difficult to effectively coordinate traffic between the two systems without first settling on a consistent schedule.

Figure 1.14. Timeline showing the schedules of two signal programs for two adjacent corridors.
In addition to signal timing anomalies, oversaturation is another problem that may affect corridor operations. This occurs when traffic exceeds the provided capacity. Figure 1.15 shows a time-space diagram illustrating this situation. In this case, the green time is constrained at Int. 3 and long queues form as a result. The dashed lines show the queue-formation shockwaves cascading backwards; note that these extend past Int. 2, a spillback situation in which vehicles are unable to use the green at Int. 2 because of a queue at the next intersection. One of the vehicle trajectories is highlighted (callout “i”); this vehicle has to stop three times before it can get through Int. 2 and 3 (callouts “ii”, “iii”, and “iv”). The inability of a green indication to serve all of the queued demand is referred to as a split failure\(^1\). The effect on corridor performance is to dramatically increase the travel time; each split failure increases the travel time by about one cycle length.

\[1\] Also called a cycle failure.

Figure 1.15. Impact of queue spillback.
1.7.3 500-ft Level: the Intersection

The next level down from the corridor is the intersection, or the 500-ft level. The preceding set of graphics illustrated how the start and end of green for coordinated movements had a profound impact on travel time through a corridor. The need to maintain the coordination schedule has to be balanced against the need to serve all of the different movements at the intersection.

Figure 1.16 explains eight-phase operation, which is a template for most signalized intersection control in the US with two crossing streets with left-turn movements. Figure 1.16a shows a plan of an intersection, with each lane labeled to indicate the phase that controls its governing signal indication. For example, phase 1 controls the protected southbound left turn; phase 2 controls the protected northbound through, the permitted northbound right, and a pedestrian movement; and so on. In practice, the phasing scheme is tailored to the particulars of an intersection and these can vary greatly.

Figure 1.16b shows a ring diagram explaining how the phases are sequenced in the control scheme and the movements which belong to each. A brief explanation of this diagram follows, which also explains basic intersection operation. Each numbered block in Figure 1.16b represents a phase, and each row of blocks is a ring. Each ring serves one phase at a time, so with two rings, the signal controller operates two phases at a time. Phases are usually compatible with another phase in another ring, such as two opposing through movements (e.g., phases 2 and 6), or an adjacent through movement and left turn (e.g., phases 1 and 6). This scheme allows the green times for some individual movements to be made more flexible. For example, two opposing lefts such as phases 1 and 5 can be served for different amounts of green time. When phase 1 serves all of its traffic, it can end immediately and the controller can move on to phase 2 rather than having to wait for phase 5 to end. This is allowable because phase 5 is compatible with both phases 1 and 2. In a typical ring diagram, groups of compatible phases are often separated by “barriers” such as those indicated by the thick vertical lines in Figure 1.16b that separate phases 1, 2, 5, and 6 from phases 3, 4, 7, and 8. In practice, many modern signal controllers use a compatibility matrix to determine what two phases can operate simultaneously, and the control scheme can support a variety of flexible configurations. Other control schemes may involve overlaps and logic functions that can modify the output displays.

The purpose of presenting the ring diagram is to illustrate how intersection capacity is apportioned to competing movements. Under coordinated operations, the *split* of each phase controls the duration of green for each individual phase. A split is the proportion of the cycle length assigned to each phase; the sum of splits for the phases for any given ring is equal to the cycle length. Looking back at Figure 1.9, the split was shown as the duration of green time for the coordinated movement at each intersection along the arterial. This is accomplished in the control scheme by allocating a portion of cycle as the split time for the appropriate phase. For example, the “northbound” movement might be assigned to phase 2, and it could be assigned a split equal to 50% of the cycle. Under non-coordinated operations, other control parameters may be used, such as the maximum times, but the principle is the same: the duration of green is the traffic control policy that, combined with the traffic demand, largely determines the capacity performance outcomes and profoundly impacts coordination.
(a) View of an intersection operated with a typical eight-phase control scheme.

(b) A ring diagram for eight-phase control.

Figure 1.16. Twelve movements at a typical four-leg intersection.
Figure 1.17 presents the basic queue theory for traffic signals [6,7]. The plot shows the cumulative number of vehicles at the intersection over time, with the thick curve representing departures and the thin curve representing arrivals, as labeled. The slope of the arrival curve is equal to the arrival rate, which is constant in this example. During a red signal, no departures occur, so the departure curve becomes horizontal, and it diverges from the arrival curve. When the signal becomes green, departures occur at the saturation flow rate, and the departure curve turns upward until it meets the arrival curve again. The vertical distance between the arrival and departure curves represents the size of the queue, and the overall area of the polygon formed between the two curves represents the total delay experienced by vehicles on this movement.

![Diagram of queue theory](image)

Figure 1.17. Theoretical framework for bulk service of queued demand with steady arrivals.
In practice, both the arrival rate and green times fluctuate. Figure 1.18 illustrates some examples of such variations that lead to a few cycles where queues do not clear during the green time.

- In Figure 1.18a, the green times during each cycle vary, with the 2nd and 3rd cycle having shorter greens and the 4th cycle having a longer green. There are still queued vehicles at the end of the 2nd and 3rd greens, with the queue finally dissipating in the 4th cycle.
- In Figure 1.18b, the arrival rate increases in the 2nd cycle, leading to a queue buildup that does not clear at the end of the 2nd and 3rd greens. The arrival rate decreases near the end of the 3rd cycle, and the queue dissipates in the 4th cycle.

Figure 1.18. Impact of fluctuations in green times and arrival rates.
The presence of unserved demand at the end of green, as in the 2nd and 3rd cycles in the examples of Figure 1.18, are split failures—the same condition as seen in Figure 1.15 in a corridor-level view. Each vehicle affected by split failure experiences an increase in its delay roughly equal to one cycle length, because the signal will need to rotate through all of the other demands before it can return to the affected movement.

Figure 1.19 illustrates the impact of a split failure on traffic in more detail. A time space diagram is shown for two cycles, with views of traffic illustrating conditions at different times. At the start of green during the first cycle, a long line of vehicles has accumulated (“i”). A few of those vehicles are remaining at the end of green (“ii”). These vehicles must wait through the subsequent red interval while the signal serves other movements (“iii”) and can finally proceed after the start of green during the next cycle (“iv”).

Figure 1.19. Illustration of a split failure.
As this discussion shows, the longer the red time, the greater the delay. Efficient operation of the intersection will avoid “wasting” capacity on movements that do not need it, so that service can be transferred to competing queues efficiently. When possible, scheduling arrivals to occur during green can also be helpful. As it involves multiple intersections, coordination is a corridor-level problem, but its effects are observed in intersection operations. While this discussion has used illustrations based on vehicle movement, the same analysis applies to pedestrians. While pedestrians do not generally form queues to cross at intersections, they experience delay corresponding to the amount of time needed for the signal to finish serving other movements.

Queue management is also an important issue affecting corridor operations. Long queues that spill outside of their storage space can cause starvation of another movement, meaning that vehicles wanting to execute the movement are prevented from doing so by the queue. Figure 1.20 shows two examples of starvation caused by adjacent lane queueing.

- In Figure 1.20a, a queue in the left turn lane queue has extended beyond its storage space, spilling out into the adjacent through lane. Vehicles that want to make the through movements are unable to make it past the left turn queue.
- Figure 1.20b shows the opposite scenario, where a queue in the through lane has blocked entry to the left turn lane. Vehicles wanting to make it to the left turn lane are unable to enter the lane.

![Figure 1.20. Illustration of starvation.](image-url)
1.8 What Can Be Measured at Each Level?

The previous discussion talked through examples of common issues related to arterial operations at the network, corridor, and intersection levels. The question now follows of whether, and how, the success of arterial operations can be measured. For example:

- Is the quality of progression sufficient?
- Is the allocation of capacity satisfactory?
- Are excessive queues occurring?
- Are pedestrians and other non-vehicle modes subject to excessive delay?

Traditionally, these questions would be answered using a handful of techniques.

- Complaint calls from the public reporting a problem with signal operations are generally an impetus for agencies to take action. However, these are often difficult to substantiate, not always reproducible, and sometimes inaccurate. Also, many problems are unlikely to generate a complaint call yet may still have a substantial impact on traffic.
- An engineer’s anecdotal experience is an important source of information for managing a corridor or intersection. However, it is unlikely that good coverage of a system, for all operational conditions, can be developed through experience alone.
- Corridor performance has traditionally been measured by the floating-car technique, in which engineers or technicians actually drive the corridor and measure their travel time, with measurement often assisted by GPS devices. This develops a detailed picture of the performance of a single vehicle on a particular route during a particular time period.
- Traffic counts are usually undertaken at regular intervals to track utilization of facilities throughout an agency’s roadway network. For signalized intersections, this includes turning movement counts, which can then be analyzed using a methodology such as that presented in the *Highway Capacity Manual* [8], or fed into a model to estimate system performance.
- Some agencies have used video monitoring to observe conditions throughout a network, which enables timing plans to be fine-tuned in response to observed conditions. While this approach can be effective, it relies on an intuitive interpretation of the video, and is difficult to scale to a large network. It also does not develop quantitative information about system performance.

The limitations of these methods are that they capture only a small portion of the overall system problem, which includes all times of day and many different routes through an arterial network. Largely manual data collection methods such as making floating-car runs or carrying out traffic counts are also labor intensive [9]. Collecting a sufficient amount of data to actively manage a system would be prohibitive by these methods. Indeed, most agencies update traffic counts on an annual basis; intervals are longer in some cases.

If one briefly considers what would be needed to develop a complete picture of performance from a corridor perspective, for example, it would likely be a view similar to the time space diagrams populated with vehicle trajectories shown in the previous discussion. To some degree, traditional manual data collection methods aim to capture some of this information. A floating-
car run obtains one sample trajectory, and manual counts obtain an estimate of vehicle volumes at different locations in the system.

Since the objective of a floating-car run is to measure travel time on a corridor, increasing the number of probes enhances the data. This is impractical to achieve by driving the corridor directly. However, vehicle re-identification techniques can produce this data in an automated fashion with far less effort. This technology matches a vehicle at two different locations in the system and tabulates a sample travel time from that data. Automating this process allows the sampling of a large number of travel times along a route, orders of magnitude greater than can be achieved by manual methods [10,11,12]. This is accomplished by probe vehicle data, as discussed in Module 2 and demonstrated in subsequent modules. Also, with 24-hour data collection, it becomes possible to establish a complete record of roadway performance over time, which cannot be done by floating car studies.

Manually-collected turning movement counts are another conventional data source relevant to arterials. These are typically recorded in 15-minute intervals, and used to estimate capacity utilization, which requires an assumption about the green times provided at the intersection [8]. Manual counts are costly to obtain and, while generally accurate on a 15-minute scale, do not provide much real detail about the quality of operations, except on a very aggregate basis when combined in a model with estimates of the signal timing.

Automation of the counting process enables the development of a record of vehicle volumes covering a continuous time period. This can be done through analysis of the vehicle detections that occur at a signalized intersection. Additionally, if the vehicle counts are recorded with the precise timestamps in which they enter a particular location, this can be directly related to the actual signal state at that time, rather than an estimates or educated guesses. These two types of basic data elements, the detections and the signal changes, represent the essence of traffic event data, which can be recorded by many different traffic signal controllers. More advanced data elements that integrate more deeply with the traffic control logic could additionally be provided to evaluate whether the control system is adequate or whether it needs to be fine-tuned to respond to traffic more effectively.

A practical example can be developed from the objective of providing smooth traffic flow along a route in a network of arterials. This can be achieved by a strategy of coordinating traffic signals. The success of the strategy could be measured in several different ways at different levels, as explored in the following discussion.
• At a network level, measurements of traffic demand and origin-destination characteristics would determine whether the appropriate routes are selected for coordination. For example, in some systems, it may not be sufficient to coordinate the signals strictly along two-directional “mainline” routes. It has to be decided where to create break-points between systems, and whether (and how) to include multiple crossing arterials into a wider coordinated system.

Figure 1.21 shows an example network-level analysis [13] based on probe vehicle travel times, in which the median travel times are plotted (along with the 75th and 25th percentiles) for different routes through an arterial network. In this case, a detour for a freeway closure is routed through the network, and four potential alternative routes are identified. The analysis of the measured travel times reveals that the “official” detour route has a substantially higher travel time than three alternatives.
At a corridor level, measurements of travel time can determine whether the current coordination strategy performs as expected. Figure 1.22 contains a plot with hundreds of individual vehicle travel times measured along a particular arterial corridor. Each one is equivalent to a single floating car travel time run. Studying the cloud of points alone reveals a substantial increase in the travel time on Saturdays, which suggests that there are opportunities to target the Saturday timing plan for improvement.

Figure 1.22. Corridor-level performance analysis: Day-to-day travel times along a corridor [2].
• At an intersection level, measurements of demand and service characteristics can identify whether there are capacity deficiencies, which would limit the ability to provide smooth traffic flow, and whether these deficiencies can be remedied. These can further be sorted by time of occurrence and severity to better target resources [14].

Figure 1.23 shows the prevalence of split failures on eight phases of a signalized intersection over a 24-hour period, showing how often leftover demand exists at the end of green for those phases. Clearly, phase 4 and phase 5 experience the most split failures, and they are distributed throughout much of the day. In contrast, the other phases experience substantially fewer phases. This suggests that there may be an opportunity to fine-tune the green time distribution, as by changing the splits, at this intersection.

There is almost always a tradeoff in intersection operations between serving demand for service on a phase as expediently as possible, and establishing smooth traffic flow on the coordinated movements. The objectives selected for operating the system set the target balance of performance between coordinated and non-coordinated movements. Performance measures such as the ones presented here would enable the analyst to gauge whether the appropriate balance has been achieved.

Figure 1.23. Intersection-level performance analysis: Split failures by movement.
1.9 Implementation and Integration of Performance Measures

The measure of success in the implementation of an arterial management system will not be in its deployment alone, but the degree to which it is actually used to support staff duties. The purpose of such a system is essentially to make it easier for agency staff to accomplish the tasks attendant to operating an arterial, and by doing so, help the agency achieve its objectives. Therefore, the concepts presented in this report need to be integrated with those duties in order to facilitate their use and to achieve benefits from the deployment of performance measures.

Personnel categories and task distributions vary widely from one agency to another. A few survey-based reports have examined these in some detail [15,16]. Table 1.3 presents a list that has been gleaned from a review of those reports, along with four basic categories of staff involved in arterial operations. Several of these tasks would be improved by deployment of data collection systems to provide performance measures similar to the ones in this report. These are discussed in further detail below:

- **System evaluation.** Most of the performance measures discussed in this report focus on evaluating aspects of system performance. The performance measure data is intended to help traffic engineers determine whether action is needed to improve arterial operations. However, this is not the only use case targeted in this report.

- **Control plan development.** Beyond evaluation, the data obtained by the arterial management system can be used to optimize the existing signal timing plans and to validate new timing strategies after they are implemented.

- **System checking.** Traffic management center (TMC) operators and other staff in similar “analyst” type roles are sometimes tasked with managing traffic control on signalized facilities in real time. Often these are driven by video feeds of live conditions. Their activities could be more effective if coupled with quantitative performance measures to validate whether an activity had a positive effect, or to see whether the observed conditions are recurring or non-recurring.

- **Complaint handling.** Complaint calls made by the public about signal operations are often challenging to validate and reproduce; the information given is sometimes vague and it is not clear whether a problem actually occurred. With a dataset for comparing against, those complaint calls can be more effectively triaged. The performance measures also would provide a way to demonstrate that the complaints have been resolved.

- **Maintenance requests.** The placement of maintenance calls by engineering staff and TMC personnel would be improved by corroborating observed problems with quantitative data. The functionality (or lack thereof) of system components would be identifiable from their impacts on traffic performance.

- **Implementation of contingencies.** Most agencies have to contend with occasional instances of exceptional traffic conditions, such as may occur with inclement weather or special events. These are challenging to manage because it is rarely possible to develop control plans ahead of time to manage these conditions. Continuous data collection would build a dataset that would enable such contingencies to be developed, and assist in the development of after-action reports to understand whether traffic control strategies were effective.
• **Detection system maintenance.** Detector faults can have substantial effects on traffic signal operations. However, they are not always easy to detect. Generally, signals will continue to cycle even with detector failures, and public complaints are less likely to be generated as a result. Monitoring of the detection data would better inform maintenance staff as to which detectors are working and which are not.

• **Communication system maintenance.** Similar to errors in detection systems, faults in communications systems would be detected through use of an arterial management system that communicates to the field.

• **Traffic equipment troubleshooting and repair.** More intelligent generation of maintenance requests would be possible with the use of performance measures, since they would reveal locations where the equipment is not working properly. In addition, the performance measures would provide a tool for maintenance technicians to validate whether the components are working properly after a fix has been made.

Table 1.3. Personnel roles in arterial management and opportunities for performance measures to assist and improve tasks, and new roles (highlighted) required for successful deployment of performance measures.

<table>
<thead>
<tr>
<th>Task</th>
<th>Traffic System Engineer</th>
<th>Traffic Analyst / TMC Operator</th>
<th>Maintenance Technician</th>
<th>IT Specialist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Management</td>
<td>Assisted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System Design</td>
<td>Assisted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System Evaluation</td>
<td>Improved</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Plan Development</td>
<td>Improved</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System Checking</td>
<td>Improved</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complaint Handling</td>
<td>Improved</td>
<td>Improved</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maintenance Requests</td>
<td>Improved</td>
<td>Improved</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implementation of Contingencies</td>
<td>Improved</td>
<td>Improved</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data System Maintenance</td>
<td></td>
<td></td>
<td>Required</td>
<td>Required</td>
</tr>
<tr>
<td>Data Curation</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
<td>Required</td>
</tr>
<tr>
<td>Traffic Equipment Installation</td>
<td></td>
<td></td>
<td>Assisted</td>
<td></td>
</tr>
<tr>
<td>Detection System Maintenance</td>
<td>Improved</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication System Maintenance</td>
<td>Improved</td>
<td>Improved</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traffic Equipment Troubleshooting and Repair</td>
<td></td>
<td></td>
<td>Improved</td>
<td></td>
</tr>
<tr>
<td>Equipment Inventorying</td>
<td>Assisted</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Accompanying the potential improved tasks, a handful of new tasks are needed to deploy a data-driven performance measurement system. One of these is data system maintenance: the installation and operation of database servers and software to house the data and make it available to analysts. While most agencies have some IT staff that maintains computers and other equipment, the need to operate a server with many transactions on a relatively large data set will be a new IT use case for some agencies.

The other new task in the list is data curation. Data curation refers to the activities needed to maintain data in the long term. Some agencies may not want to maintain data beyond a certain horizon, whereas others might want to archive it for future uses. While ultimately the means to do that will depend on decisions of traffic engineering staff and implementation by IT staff as the use of the data increases, the users become stakeholders in the data curation process. The simplest example would be a situation where data is used for a variety of day-to-day operational purposes, and to support some before-after comparisons. However, later on, that data might be desirable to also use for a planning application. The data set must be properly curated in order for multiple groups to make use of it.

1.10 Summary

This module discussed the role of arterials as a part of an overall transportation system, and the state of arterial management systems. The importance of setting agency objectives was discussed, and various potential objectives were examined. The concept of analysis levels corresponding to the network, corridor, and intersection were introduced, to understand how objectives can be interpreted from these different spatial perspectives.

Various issues specific to the mobility performance of arterial operations were then mentioned for each of the three levels.

- Network-level performance measures consist of aggregations from smaller network components (corridors and arterials), and the origin-destination patterns throughout a system.
- Corridor-level performance measures focus on traffic flow along routes in a network. The impact of poor intersection configurations on corridor performance was illustrated through a few examples.
- Intersection-level performance measures focus on the allocation of capacity on competing movements at an intersection. Problems related to insufficient capacity include split failures and starvation.

Finally, the potential for new data sets to measure aspects of performance at the network, corridor, and intersection levels was discussed. Aspects of performance related to corridor and intersection objectives were identified and presented through examples that foreshadowed the use of data sets that are introduced in subsequent modules of the report: probe vehicle travel time and high-resolution traffic event data.
1.11 References

Module 2. Introduction to New Technologies

2.1 Introduction

This module introduces two data sets for measuring the performance of intersections and arterial corridors and networks. The first is vehicle re-identification data, and the second is high-resolution traffic event data. This module offers an introduction for readers who are unfamiliar with either or both of these two data sets. Included are some example applications showing ways that each data set can be put to use. Later modules will delve into applications in greater detail.

2.2 Vehicle Re-Identification Data

2.2.1 Vehicle Re-Identification and Travel Time Measurement

One critical aspect of the performance of any transportation network is the amount of time required to get from an origin to a destination. Measuring this can be done by tracking travelers through the network. For a roadway network, this was historically accomplished with manual license plate matching studies. By identifying the time a vehicle arrives at two different locations, the travel time for that vehicle is determined by taking the difference of the observation times. This is vehicle re-identification [1].

Floating car studies are another way of measuring travel times. The analyst actually drives a vehicle along a route, and records the time that it takes to traverse the system. Because the total travel time of the probe vehicle is determined by the start and end times, each floating car run yields one travel time data point.

Figure 2.1 illustrates how travel time and delay would be calculated for vehicles traveling on an arterial using vehicle re-identification data. This plot is a time space diagram, with time on the horizontal axis and distance on the vertical axis. The lines represent the trajectories of vehicles as they move along the roadway; two trajectories are highlighted (callouts “i” and “ii”). The red bars represent times that the signal at each intersection was red, and vehicles must stop.

Along the vertical axis, callouts “iii” and “iv” represent the locations of two checkpoints where the unique signature of a vehicle is recorded. The time when each vehicle crosses that point is written down as an observation. The difference between the observation times gives the travel time, as indicated by the intervals at the top of the chart. For each vehicle, the dashed blue line represents the ideal travel trajectory, in absence of any speed reductions. It begins at the point when the vehicle crosses the first checkpoint. Conceptually, the delay is represented by the time difference between when the dashed line crosses the second checkpoint, and the actual time that the vehicle crosses. This is equivalent to subtracting the ideal travel time from the actual travel time to obtain the delay.
This analysis allows variations in performance of an arterial corridor to be measured. As shown in Figure 2.1, the same signal operation can lead to some vehicles experiencing a rather low amount of delay, such as the first vehicle (callout “i”), while others in the same stream may experience more delay, such as the second (callout “ii”). To develop a good understanding of the performance, a large sample of travel times must be collected in order to assess the central tendency and the amount of variation, or the reliability of the travel time.

Figure 2.1. Travel times from vehicle re-identification.
2.2.2 Technologies for Automated Vehicle Re-Identification

In recent years, a number of emerging technologies have enabled travel times to be measured on highways [1]. However, the analysis of travel time characteristics for signalized facilities is particularly challenging, because of the considerable variability in travel times. This variability results from the complex, dynamically changing nature of signal control and traffic patterns along a surface street corridor having many entries and exits. Because of the growing emphasis on performance measures (most recently in the MAP-21 transportation bill), there is very strong demand for arterial performance data, but providing reliable information for signalized facilities has proven to be a demanding endeavor in comparison to generating similar measures for limited access freeways.

Automation of the vehicle re-identification process enables a sufficient number of samples to be collected to enable meaningful statistical comparisons over a much longer time period and at a much lower investment of resources than by manual methods. Automating the process also makes it more of a random sample, mitigating potential problems of bias that might be associated with the driving behavior of the floating car driver.

Automated vehicle re-identification data is developed by sensing a “signature” of a vehicle at two or more locations. Various technologies have been applied to accomplish this, but one of the most successful has been with the detection of Bluetooth media access control (MAC) addresses [2,3,4,5,6,7]. A MAC address is a hexadecimal string used by network-capable electronics that provides an identifier for a device in many different applications. The 48-character MAC address provides for $256^6$ different hexadecimal strings (about 281 trillion). Thus, while multiple devices technically can have the same MAC address, it is unlikely, so the MAC address is effectively a unique identifier. Devices such as mobile phones broadcast this information when they are wirelessly connected to other devices. Bluetooth is a standard protocol used for that purpose. For example, a driver using a wireless headset to communicate with a mobile phone will likely be broadcasting a MAC address, which can be read by any other Bluetooth antenna. This allows a unique identifier to be collected at the sensor location. If that driver is observed at another location, the travel time of the vehicle can be deduced.

This technology has been adopted by several different vendors for providing traffic data. In addition to Bluetooth MAC addresses, there are other broadcasted identifiers that can be observed and used with the same analysis, such as Wi-Fi identifiers.
2.2.3 Examples of Vehicle Re-Identification Travel Time Data

Figure 2.2 shows an example of an application of vehicle re-identification data. This study focused on a spike in traffic demand due to a special event, and the impact on performance of an arterial route that received a substantial amount of that added traffic. Four sensors were stationed at various locations along the route, as indicated in the map. Three matching pairs are identified and the travel times for each of these are plotted in the chart; each point represents a matched vehicle from the starting point of the arterial section, to three successive endpoints. Additionally, three points in the chart represent a floating car run undertaken at the same time.

The use of automated vehicle re-identification enabled dozens of data points to be collected over a three-hour period, whereas the floating car yielded a single data point. Consequently, the automated data captured a trend that occurred during that time period whereas the single floating car could not do so. In addition, it can be seen that the floating car drove along the corridor somewhat faster than other traffic; its travel times are considerably lower than most of the other observations. In other words, the floating car did not accurately capture the trend in travel times during the same time period.

Figure 2.2. Example data from an automated vehicle re-identification travel time study [3,8].
This study used temporary sensors to monitor a special event. Permanent sensors provide coverage over a much longer time scale, allowing an even larger sample size to be obtained. Figure 2.3 shows a map of an arterial, Maryland State Route 355, where permanent sensors were installed. The locations of signalized intersections are indicated by the red markers while the blue markers indicate two data collection devices at two arterial endpoints.

Figure 2.4, showing the raw northbound travel times collected over a 6-month period, includes approximately 50,000 observations. One can visually distinguish the densest range of the point cloud (between 5–10 minutes), but the range of observations extends as high as 20 minutes or more. Many of those longer observations are probably representative of travelers who made a stop while moving along the route. Apart from these prominent features, it is difficult to further determine whether trends exist within this data—such as whether the distribution becomes heavier in one range of travel times or another, or how the system performance varies between daily timing plan intervals.

To begin extracting useful performance measures from this complex portrait of the travel time (50,000 samples in this case), a first step is partitioning the point cloud into distributions that can be aggregated. For a signal system, an appropriate template for dividing up the daily timeline is the programmed timing plan schedule. Figure 2.5 shows the northbound travel time observations on Wednesday, 10/26/2011. The signal time-of-day schedule change patterns at 06:00, 09:30, 15:00, 19:00, and 22:00, are indicated by the vertical dashed lines.

Figure 2.6 shows the distribution of travel times observed during the AM timing plan (06:00-09:30) throughout the sample data, which includes approximately 5,000 samples. This is represented as a histogram and as a cumulative frequency distribution (CFD). On top of this, the 25th percentile, median, and 75th percentile travel times are shown. These numbers indicate that a typical northbound travel time through the system required about 9 minutes, and that 50% of the travel times fell between 7.5 and 16.5 minutes. The difference between the 75th and 25th percentiles, called the interquartile range, conveniently measures the degree of variability in the travel time, which relates to the reliability of the system.

The CFD ends at a value of 45 minutes for the 90th percentile (0.9), which means that about 10% of the observations had longer travel times. These observations most likely represent vehicles that have intermediate destinations along their northbound journey. Filtering of the raw data will exclude these vehicles from the analysis to avoid inflating the travel time.
Figure 2.3. Test corridor for vehicle re-identification travel times.
Figure 2.4. Raw travel time data for northbound vehicles from a 6-month time period.

Figure 2.5. 24 hours of raw data for northbound vehicles on Wednesday, 10/26/2011.

Figure 2.6. Distribution of travel times for northbound vehicles during the AM peak (6:00-9:30).
Figure 2.7a presents the CFDs for northbound movement for all of the Wednesday mornings in the data set. In this case, a 30-minute filter has been applied, thus setting the 100th percentile of all the CFDs at 30 minutes. This allows the individual CFDs to be compared without having to consider variations in the percent of vehicles spending time at intermediate destinations. Over the 6-month period, the median value hovers around 9 minutes, and the interquartile range tends to vary from about 7.5 to as high as 15 minutes for most days. A few outliers are apparent: two days with shorter travel times (callouts “i” and “ii”) and two with longer travel times (callouts “iii” and “iv”). Figure 2.7b presents the same information as a “box-whisker” plot, in which the 25th percentile, median (50th percentile), and 75th percentile are indicated. This allows the data to be visualized summarily over a long time period.

Figure 2.7. Identifying outliers in arterial travel time data.
2.2.4 Applications of Vehicle Re-Identification Travel Time Data

A couple example uses of vehicle re-identification data are briefly presented here, with more details to follow in subsequent modules. Perhaps the most obvious application of travel time data is to compare performance under different conditions. Figure 2.7 illustrated fluctuations in travel times occurring on the same corridor over a 6-month period. The same type of data can be used in a shorter span of time under more similar traffic and weather conditions to examine the impacts of operational changes.

Figure 2.8 presents two CFDs of travel time along a corridor showing the impact of an improvement in signal timing. The two directions of travel along the arterial (northbound and southbound) are shown in the two CFDs. As indicated by the distance between the two curves at the median or 50th percentile, typical corridor travel times were improved by over a minute in both directions. This type of metric demonstrates a return on investment in adjusting the traffic control, or to validate whether an improvement has had a positive or negative impact, or whether it has achieved a target level of performance.

The same analysis can be carried out on the network level to compare multiple routes, as demonstrated by the comparison of alternate routes in Figure 2.9. Four box-whisker plots are presented. These show the performance of four alternative routes between the same origin and destination in an arterial network under an increased traffic load caused by the closure of an area freeway. The performance is partitioned between weekend and weekday performance. In addition to the observed travel times, the proportion of observed travelers using each route is also included. The results reveal that only a small number of motorists use the official detour route, while other routes generally have better performance. Such information could be used to evaluate work zone and/or detour management in an after-action report, or for planning purposes in evaluating origin-destination characteristics through a network.
Figure 2.8. Travel time cumulative frequency distributions before and after optimization of signal timing [9].

Figure 2.9. Comparison of performance and proportion of observed vehicles on alternative routes through an arterial network [10].
2.2.5 Comparison to Crowdsourced Probe Data

In addition to vehicle re-identification data, some commercial data providers have brought data to market based on aggregated speeds of probe vehicles, whose positions and speeds are collected using GPS measurements obtained by navigational devices and mobile phone applications.

While this data set allows the development of data over a large geographic extent without the use of any field infrastructure, the data itself differs substantially from vehicle re-identification data. Instead of individual vehicle travel times, the data consists of aggregated speeds on a minute-by-minute basis per roadway segment. Consequently, the speeds are converted into travel times and aggregate across roadway segments to develop a travel time metric. The result is an estimate of travel time. It is not necessarily the same vehicles reporting speeds in neighboring segments, so the accuracy of the estimate is unclear. Also, the resolution of the data is controlled by the segment endpoints, the 1-minute time aggregation, and the units and fidelity at which speeds are reported.

While the data provides a reasonable estimate of arterial performance for high traffic arterials with a relatively low density of traffic signals [11], some direct comparisons against vehicle re-identification data have shown a considerable discrepancy between the two [12]. While more such comparisons are needed to draw a conclusion about the relative quality of crowdsourced data, at this time, crowdsourced data may be regarded as providing a rough estimate of performance that can be compared within the same data set to track performance over time or between similar facilities. However, it does not provide actual travel times of individual vehicles directly.
2.3 **High Resolution Traffic Event Data**

2.3.1 **Traffic Control Development Process**

Figure 2.10 illustrates the process of developing an arterial traffic control system [13]. The first step consists of defining objectives, such as establishing the limits of the corridor, whether to coordinate the traffic signals, and so forth. The second step consists of data collection, which traditionally consists of manual turning movement counts and gathering other relevant information. That data feeds a model of the system in step 3, which creates a set of timing plans in step 4. The initial design process is represented by feedback loop 1, an iterative process through which the designer makes incremental improvements on the timing plan.

After accepting the design, field implementation follows in step 5. This consists of programming the control equipment with the timing plan. In step 6, the system performance is monitored, and feedback loop 2 consists of making incremental improvements to the timing plan, because the model will usually differ from the real world. This second feedback loop depends on the ability to make field observations. It is not possible to sustain manual observations very long, so eventually this round of feedback ceases, in absence of data collection. Thus, as changes to the system accumulate over time, the control plan is not updated except at very infrequent, arbitrary retiming intervals, or in response to problems that have generated a user complaint.

Feedback loop 3 represents the process by which the results of system monitoring are used to feed back into the planning and design stages. This use case is not supported by current manual field observations of the operation.

By establishing a record of the activity taking place at a signalized intersection, including vehicle movements and service times, high-resolution traffic event data aims to accelerate the activity on feedback loop 2, and enable feedback loop 3 to take place. By deriving a set of performance measures that are relevant to basic traffic engineering concepts, this data set can improve the transparency and accountability of the agency operating the system in addition to affording a greater potential to improve how the system operates.

![Figure 2.10. The traffic control development process [13].](image-url)
2.3.2 High-Resolution Data Concept

Traffic control systems have been able to collect various types of data for some time. However, because of limitations in computational power and bandwidth, this data mostly consists of aggregated reports, such as detector occupancy per 15-minute bin. Thus, this form of data might be considered “low resolution” data. Some advanced control systems internally gather more detailed information, but this is primarily used to drive a responsive or adaptive decision, and the engineer cannot always easily access that data. Furthermore, each system has its own unique, proprietary method of recording and storing the data.

High-resolution traffic event data consists of a record of all measurable events affecting the control policy at a traffic signal, at the highest possible time resolution. Current signal controllers operate at a resolution of 0.1 seconds, and high-resolution data is recorded at that interval. While preceding research used external systems to obtain this data by reading directly from the electrical outputs in a cabinet, or by adding a widget to “sniff” cabinet device communications, many signal controller vendors now provide a data logger in the controller itself capable of logging the data. This reduces the expense of implementation, since some type of controller always exists at a signalized intersection, the data logger is implemented in software, and no additional device is required. Furthermore, by using the clock time internal to the controller, the problem of clock synchronization between the control system and the measurement device is avoided.

Several years of collective discussion between vendors led to the development of a specification for the data enumerations [14], which is now used by at least five different vendors in the implementation of a data logger. Because of this, the same data objects can be retrieved from different equipment, without the need to translate from one data format to another at the user application level. Instead, each vendor independently develops its own proprietary data logger and internal data format, but provides the data externally using the same format as the other vendors. This makes it possible to develop an analysis system that can handle equipment from multiple controller manufacturers.

2.3.3 Example High-Resolution Traffic Event Data

The data itself consists primarily of two basic types of information [15]:

- A record of vehicle detections, which describes the times when a detector became occupied or when it became unoccupied. This measures the behavior of traffic at the intersection or on approach to the intersection.
- A record of signal output changes (e.g., when a signal head changes to green, yellow, or red). This measures the allocation of capacity and the schedule of green times.

Legacy data collection systems generally focused either on aggregate measures of detector occupancy or green times. By combining these two types of information together, it becomes possible to measure the utilization of the capacity, and to characterize the quality of progression.

In addition to these types of information, additional control events can be logged, which would not necessarily be observable from outside of the control system. This is a further advantage to
having a data logger internal to the controller. Examples include times when control patterns change or are adjusted, when calls for service occur (which are not always directly tied to detections), and so forth. This enables a record of the control policy to be established, which further enhances the transparency of the operation. With this type of data, it becomes possible to establish the “why” behind every single decision made by the control system, which additionally helps the agency demonstrate accountability in how it allocates the right-of-way at intersections.

Table 2.1 shows an example of raw high-resolution event data. Each row represents a single event. Each event is described by three elements:

- The timestamp states when the controller logged the event.
- The event code ID represents the category of event.
- The parameter provides contextual information about the event which varies by category. For example, detection events are related to a particular detector number, while signal status (phase and overlap) events are related to a phase or overlap number.

In addition, a signal ID is used to attach the event to a particular controller and/or intersection. This is generally added at the system level when integrating the event into a database.

The events in Table 2.1 span the beginning of a call for preempt to the beginning of track clearance green. The first event occurs when the controller receives a call for preempt 2. It immediately drops coordination. About one second later, the controller terminates the current phases (2 and 6) and places omits on all of the other phases and pedestrian phases, with exception of Phase 10. This is a special phase used to implement the track clearance green through an overlap, in this particular intersection’s configuration. At 15:22:17, Phase 10 is activated, which also brings up Overlap 2, and is also registered as the beginning of track clearance green.

This is one example of how the data itself can be used to verify that the controller performs as expected. Many other additional applications can be devised, to match agency objectives. Pedestrian service, delayed detector input, detector errors, manual control via the police button, traffic responsive pattern selection, adaptive timing adjustments, and transit signal priority are a few additional use cases for which the record of events would be useful for verifying operation of the control system.
Table 2.1. Example of high resolution signal data.

<table>
<thead>
<tr>
<th>Signal ID</th>
<th>Event Timestamp</th>
<th>Event Code ID</th>
<th>Event Param</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:10.3</td>
<td>102</td>
<td>2</td>
<td>Preempt 2 Call On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:10.3</td>
<td>150</td>
<td>0</td>
<td>Coord cycle state change (Free)</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:10.3</td>
<td>0</td>
<td>2</td>
<td>Phase 2 On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:10.4</td>
<td>82</td>
<td>63</td>
<td>Detector 63 On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:10.4</td>
<td>174</td>
<td>64</td>
<td>Unit Alarm Status 1 change to 64</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>7</td>
<td>2</td>
<td>Phase 2 Green Termination</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>7</td>
<td>6</td>
<td>Phase 6 Green Termination</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>8</td>
<td>2</td>
<td>Phase 2 Begin Yellow Clearance</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>8</td>
<td>6</td>
<td>Phase 6 Begin Yellow Clearance</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>2</td>
<td>10</td>
<td>Phase 10 Check</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>46</td>
<td>1</td>
<td>Phase 1 Omit On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>46</td>
<td>2</td>
<td>Phase 2 Omit On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>46</td>
<td>3</td>
<td>Phase 3 Omit On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>46</td>
<td>4</td>
<td>Phase 4 Omit On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>46</td>
<td>5</td>
<td>Phase 5 Omit On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>46</td>
<td>6</td>
<td>Phase 6 Omit On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>46</td>
<td>7</td>
<td>Phase 7 Omit On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>46</td>
<td>8</td>
<td>Phase 8 Omit On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>48</td>
<td>1</td>
<td>Phase 1 Pedestrian Omit On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>48</td>
<td>2</td>
<td>Phase 2 Pedestrian Omit On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>48</td>
<td>3</td>
<td>Phase 3 Pedestrian Omit On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>48</td>
<td>4</td>
<td>Phase 4 Pedestrian Omit On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>48</td>
<td>5</td>
<td>Phase 5 Pedestrian Omit On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>48</td>
<td>6</td>
<td>Phase 6 Pedestrian Omit On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>48</td>
<td>7</td>
<td>Phase 7 Pedestrian Omit On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>48</td>
<td>8</td>
<td>Phase 8 Pedestrian Omit On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>48</td>
<td>10</td>
<td>Phase 10 Pedestrian Omit On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.3</td>
<td>105</td>
<td>2</td>
<td>Preempt 2 Entry Started</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.4</td>
<td>43</td>
<td>10</td>
<td>Phase 10 Call Registered</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:11.4</td>
<td>43</td>
<td>6</td>
<td>Phase 6 Call Registered</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:12.2</td>
<td>43</td>
<td>2</td>
<td>Phase 2 Call Registered</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:12.2</td>
<td>82</td>
<td>5</td>
<td>Detector 5 On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:12.5</td>
<td>81</td>
<td>5</td>
<td>Detector 5 Off</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:12.5</td>
<td>44</td>
<td>2</td>
<td>Phase 2 Call Dropped</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:15.6</td>
<td>10</td>
<td>6</td>
<td>Phase 6 Begin Red Clearance</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:15.6</td>
<td>10</td>
<td>2</td>
<td>Phase 2 Begin Red Clearance</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:15.6</td>
<td>9</td>
<td>2</td>
<td>Phase 2 End Yellow Clearance</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:15.6</td>
<td>9</td>
<td>6</td>
<td>Phase 6 End Yellow Clearance</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:17.0</td>
<td>61</td>
<td>2</td>
<td>Overlap 2 Begin Green</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:17.0</td>
<td>0</td>
<td>10</td>
<td>Phase 10 On</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:17.0</td>
<td>11</td>
<td>2</td>
<td>Phase 2 End Red Clearance</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:17.0</td>
<td>11</td>
<td>6</td>
<td>Phase 6 End Red Clearance</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:17.0</td>
<td>1</td>
<td>10</td>
<td>Phase 10 Begin Green</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:17.0</td>
<td>12</td>
<td>2</td>
<td>Phase 2 Inactive</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:17.0</td>
<td>12</td>
<td>6</td>
<td>Phase 6 Inactive</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:17.0</td>
<td>31</td>
<td>1</td>
<td>Barrier 1 Termination</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:17.0</td>
<td>61</td>
<td>1</td>
<td>Overlap 1 Begin Green</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:17.1</td>
<td>44</td>
<td>10</td>
<td>Phase 10 Call Dropped</td>
</tr>
<tr>
<td>4435</td>
<td>2015/02/16 15:22:17.1</td>
<td>106</td>
<td>2</td>
<td>Preemption 2 Begin Track Clearance</td>
</tr>
</tbody>
</table>
Examining the raw data is one use for detailed troubleshooting, but the data is more powerful when distilled into performance measures that can track the signal operation over time. One simple example of such a performance measure is effective cycle length, the amount of time required by the signal to rotate through all of its demands for service. Figure 2.11 shows three different plots of cycle length over 24 hours for different intersections under a variety of signal timing options.

- Figure 2.11a shows a typical example of a coordinated intersection. During the time period from midnight to 6:00, the intersection runs free and often dwells in green for long periods of time on the major street, leading to a wide variation in cycle length. However, starting at 6:00, the cycle length is the same during every cycle. The plateaus represent different cycle lengths occurring during different time of day plans, and the troughs between them represent one cycle of transition between plans. The intersection is coordinated through the end of the day; in the last few hours, the demand for side street movements diminishes and the intersection begins double cycling.

- Figure 2.11b shows a neighboring intersection that operates with the same time of day plan, but with an advanced controller feature called “early yield” active. This feature allows vary the starting and ending times of the coordinated green. The cycle length is allowed to vary up to about 10% of the cycle in this scheme, as can be seen by the fluctuations around a base value in each time period.

- Figure 2.11c shows an intersection that is non-coordinated through most of the day. This intersection is located in a relatively low volume corridor. From 6:00 to 9:00, a coordinated plan operates, but during the rest of the day, there is no coordination plan. The cycle length can be seen to expand and contract in response to the demand for traffic. As in the other intersections, during the overnight hours, the signal dwells in green on the major street for long periods of time, leading to very long effective cycle lengths.

This plot presents a simple piece of information, but it allows useful inferences to be made about the operation at each signal. The time-of-day plan is confirmed to operate the correct cycle length throughout the day; errors or mismatches of the cycle length could be discerned from an analysis of this data. This example only scratches the surface of what can be accomplished by reviewing the record of signal operation.
(a) A location with fixed cycle length.

(b) A location with early yield.

(c) A location under free operation for most of the day.

Figure 2.11. 24-hour plots of effective cycle length.
2.3.4 Applications of High-Resolution Traffic Event Data

Two aspects of intersection operations affecting performance of vehicles are the allocation of capacity and the quality of progression. Good local control is characterized by efficient and equitable service of competing movements, while good system control means that smooth traffic flow through a group of intersections is achieved by well scheduled signal change times. These are two common objectives in signal operation, which can sometimes compete with each other. The degree of success in either endeavor can be measured using high resolution data. The following discussion presents some example performance measures for that purpose.

Figure 2.12 presents a time-space diagram showing the concept behind calculating the volume-to-capacity ratio, which measures how much of the provided capacity is utilized. As before, the red bars represent red times on the approach, while the black lines show vehicle trajectories. In this figure, the horizontal blue lines indicate locations of detectors. In this case, an advance detector is located approximately 400 ft upstream of the stop bar, while another is at the stop bar. Vehicle detections are represented by the dots at each detector location. A backwards projection from the start of red times is used to count vehicles served during a particular cycle.

In this case, there are 12 vehicles served during the cycle that starts around 340 seconds and ends at 390 seconds. The duration of green during this same cycle is 28 seconds. If the saturation flow rate is 1800 vehicles per hour, this equates to 0.5 vehicles per second. This represents the theoretical maximum number of vehicles that could be served per second of green. From this, the volume-to-capacity (v/c) ratio is calculated from

\[
\frac{v}{c} = \frac{\text{volume}}{\text{capacity}} = \frac{N}{g \times s},
\]

where \( N \) is the number of vehicles, \( g \) is the green time, and \( s \) is the saturation flow rate in vehicles per second. Given \( N = 12 \), \( g = 28 \), and \( s = 0.5 \), this is equal to \( 12 / (28 \times 0.5) = 0.86 \) during this particular cycle.

Figure 2.13 shows a tabulation of the cycle-by-cycle v/c ratio over 24 hours for a signalized approach in the real world. This represents the performance of a mainline through movement. There is a characteristic peak in the AM period (callout “i”), during which the volume regularly exceeds the capacity (that is, it rises above 1.0). For the rest of the day, the volume is much lower. Split failures are probably occurring, with the movement unable to serve all of the demand during one cycle. The rest of the day does not come close to this peak. After 19:00, the v/c ratio levels off (callout “ii”) and becomes quite minimal during the overnight period when the signal dwells in green on the mainline, making the provided capacity on the movement much higher than the volume.

Based on that analysis, one might conclude that this movement should receive more green time during the AM peak. However, this would require taking time away from competing movements to accomplish it. An analysis of the v/c ratios (or other utilization metrics) for those competing movements would be helpful to evaluate whether this is feasible. This and other examples will be explored further in Module 4.
Figure 2.12. Volume-to-capacity ratio calculation concept.

Figure 2.13. Volume-to-capacity ratio over 24 hours [13].
Figure 2.14 illustrates how a similar approach can be taken to measure the quality of progression by relating detection activity to green times. Here, the occurrences of red and green at the intersection are projected backward to the times when vehicles are observed at the advance detector. This determines whether the vehicles arrive on red or on green. From this, the total percent on green (POG) can be calculated.

\[ POG = \frac{\text{arrivals on green}}{\text{total arrivals}} = \frac{N_g}{N_g + N_r}, \]  

Equation 2

Here, \(N_g\) is the count of arrivals on green and \(N_r\) is the count of arrivals on red. In Figure 2.14, there are five arrivals on red and six arrivals on green, so the POG is \(6 / 11 = 54\%\).

A 24-hour view of this performance measure at a real world location is presented in Figure 2.15. The performance is shown for a coordinated movement with an advance detector. The POG reaches the highest values during the 9:00–12:00 time period (callout “i”), with up to 90\% of the arriving vehicles seeing a green indication at the time they were detected. The POG falls to between 60\% and 70\% during the PM peak around 15:00 (callout “ii”). Other time periods fall somewhere in between. During the low-volume overnight period, which includes the hours after approximately 19:00 and before 6:00, there are many cycles with one or two arrivals, leading to several 100\%, 50\%, and 0\% values.
Figure 2.14. Percent on green calculation concept.

Figure 2.15. Percent on Green over 24 hours [13].
By compiling composite performance of individual movements at the intersection and network level, it becomes possible to identify “hot spots” in the network and target resources to proactively improve deficiencies. Performance measures such as v/c ratio and POG would respectively allow the capacity and progression performance of the system to be characterized.

Figure 2.16 shows an example where all of the intersections in a system are ranked from most to least severe rate of split failures. The data is partitioned to investigate how network performance varies according to time of day. Unsurprisingly, the PM peak tends to have a high rate of failure throughout the system. However, perhaps unexpectedly, the midday performance includes a number of failures, despite overall volume levels being somewhat lower than during the peaks. The AM peak, on the other hand, has a rather low amount of failures. On reflection, considering that the midday time period likely has a higher demand for crossing street traffic than other times of day, while timing plans still generally favor “mainline” movements at the expense of cross-street traffic, the outcome is perhaps less unexpected. However, the data seems likely to challenge preconceived notions about system performance.

This is an example of how performance data can identify opportunities for improvement across a large deployed system. The performance measures presented here are only a preliminary example; subsequent modules explore more in-depth uses of the data to reveal additional aspects of performance with greater detail.
2.4 Summary

This module provided an overview of two data sets with applications to performance monitoring of arterials: vehicle re-identification data, and high-resolution traffic event data.

- Vehicle re-identification data is developed by matching unique identifiers of vehicles at multiple points through a roadway network. This enables an analyst to characterize the performance of corridors and other routes and make operational comparisons. Examples were presented for comparing performance before and after a timing change, and for comparing different routes from a common origin and destination.

- High-resolution traffic event data is generated by traffic signal controllers which record timestamps of vehicle detections, signal state changes, and other information. These are recorded to the nearest tenth of a second, which is the smallest time interval that current signal controllers register for purposes of operating the control logic. The data allows detailed diagnostics, but can also be aggregated into performance measures from the movement and intersection up to the network level.

Subsequent modules of this report explore the applications of these technologies in more detail at the intersection, corridor, and network levels. Module 3 investigates travel time and travel time reliability, and Module 4 looks at capacity and demand.

2.5 References


11. Young, S. “Use of Probe and Bluetooth Data for Arterial Performance Measures in the I-95 Corridor Coalition.” Presented at the NATMEC Workshop on Arterial Data and Performance Monitoring, Chicago, IL, June 2014.


Module 3. Travel Time and Travel Time Reliability

3.1 Introduction

This module approaches arterial performance from the perspective of the user by assessing performance based on observed travel times and the reliability of those travel times. These performance measures enable the analyst to characterize the central tendency of travel times, and whether they are consistent from day to day and hour to hour. In addition to directly measuring travel times, a detailed analysis of traffic signal operations along the route can be developed by measuring temporal relationships between traffic flows and green times at intersections.

The focus of this module is the effectiveness of system control in an arterial network, and the aspects of that control as it appears at the intersection, corridor, and network levels. One aspect of arterial performance central to this discussion is the relationship between intersections, and how they perform individually and collectively: the progression of traffic through the system. This relates to common operational objectives such as the provision of smooth traffic flow and of minimizing delays for roadway users. There is some interplay of progression with the important issue of capacity utilization, which is discussed further in Module 4.
3.2 Intersection Level Analysis

The travel time experienced by a vehicle making its way through an arterial network can be considered as the result of a series of trips from one intersection to the next. The intersections are the primary locations where delay is accumulated during the overall trip, and the amount of delay incurred at an individual intersection is determined chiefly by whether the vehicle arrives during a green indication, or whether it has to stop at a red signal or at the back of a queue.

Under completely ideal conditions, each vehicle would arrive during a green indication and progress through each intersection without having to re-adjust its speed. However, the arriving traffic stream includes vehicles from multiple upstream movements. In addition to platoons of “mainline” vehicles, there are also entries from side street left and right turns as well as midblock entries from driveways. The only way to serve all of this traffic would be for the signal to never turn red. Instead, some of the traffic must be prioritized over the rest. Complicating the situation further, each intersection has a varying level of demand for competing phases, so a series of intersections will likely have different amounts of green time that can be allotted to coordinated movements.

A starting point for establishing a target for intersection performance is the following (paraphrased) rule of thumb: “Try not to stop vehicles, but if they must be stopped, don’t stop them for very long.” [1]. This adage aligns with objectives aimed at achieving smooth traffic flows and reduced travel times along arterial routes (including both coordinated and non-coordinated movements at intersections) and contains two measurable elements:

- The arrival characteristics at an intersection when the upstream inflow can be measured (i.e., whether vehicles arrive on green or on red); and
- The delay times associated with individual movements at an intersection.

At the corridor level, two groups of performance measures are developed that examine these two aspects of operation in detail.

3.2.1 Traffic Flow through an Intersection

The arrival characteristics of arrivals at an intersection are important to the travel time characteristics of routes including the associated movements at that intersection. Vehicles that arrive on red must stop while those that arrive on green are able to proceed through the intersection without stopping (allowing, of course, for queues to clear). Since vehicle arrival times and signal states can both be measured, each arrival can be determined as taking place during red or green.

Figure 3.1 introduces a graphical tool called the “Purdue Coordination Diagram” (PCD) [2,3] for visualizing arrivals that superimposes a record of vehicle arrivals onto a representation of the red and green times. Figure 3.1a shows a time-space diagram with a series of vehicle trajectories representing arrivals at the advance detector. The signal state at the time of arrival is also shown by vertical demarcations. In Figure 3.1b, the same events are represented for a single cycle in a chart. The horizontal axis represents the time of day, while the vertical axis represents the time in
cycle. The detection times for each vehicle are plotted using the timestamp as represented by both time of day and time in cycle. The vertical space is segmented into red, green, and clearance times according to the time during the cycle in which these occurred. Vehicles 1–8 arrive during red, and vehicles 9–13 arrive during green. Figure 3.1b represents a PCD for a single cycle.

The figure becomes more powerful when examining multiple cycles in succession. Figure 3.2 shows a PCD for a series of cycles spanning a 30-minute period. As before, the horizontal axis represents the time of day that each event occurs in, while the vertical axis shows the time in cycle. Vertical lines demarcate the end of each cycle. Moving from bottom to top, the cycle begins at the previous beginning of red. The next signal event is the beginning of green during this cycle (indicated by the green line). Detections occurring after this event represent arrivals in green, while detections occurring before this event represent arrivals in red. Thus, dots above the green line are characteristic of good progression, while those falling below the green line represent poor progression. The last signal event is the end of green (indicated by the red line), which marks the end of the cycle and the beginning of the next.

Clusters of dots represent platoons of vehicles. In Figure 3.2, one can discern large platoons (such as callout “i”) that likely originate from the upstream major movement. This can be considered the “primary” platoon that would usually be optimized for arrivals on green. Some smaller “secondary” platoons that originate from upstream turning movements (such as callout “ii”) are also apparent. In this case, because the green time represents a large portion of the cycle, both groups of vehicles usually arrive on green.

Figure 3.3 expands the PCD to a view of 24 hours of operation. Included is an additional plot of the percent on green (POG) for comparison. As in Figure 3.2, the green line indicates the beginning of green, the red line indicates the end of green, and the dots show vehicle arrival times. The vertical blue lines represent the times when the time of day (TOD) pattern switched from one to another. The chart of POG shows that for most of the day, the timing plan delivers most of the incoming traffic flow to a green indication, with about 90% of the arrivals occurring during green. Pattern 3, in operation around 13:30–15:00, has a slightly lower POG that corresponds to some of the secondary platoon in the PCD arriving just prior to the start of green while in other patterns this group of vehicles arrives within the green.

A glance at this PCD would reveal that an objective of providing good quality progression is met fairly well during most of the day, with minor opportunities to improve pattern 3. The PCD also shows that coordination begins at 5:00 and lasts until the end of the day, and uses a cycle length of 140 seconds for this entire time period. However, during the early morning and late evening, there are significant portions of time in which there are no arrivals. There are probably opportunities to consider using a lower cycle length during these shoulder periods.
(a) Record of vehicle arrivals over time and hypothetical vehicle trajectories.  
(b) PCD for a single cycle.

Figure 3.1. Purdue Coordination Diagram (PCD) concept [2].

Figure 3.2. PCD depiction of arrivals over several cycles.
Figure 3.3. 24-hour Purdue Coordination Diagram.
A variety of different conditions can be identified using PCDs, as illustrated by the following graphs, which are paired conceptual time-space diagrams.

- Figure 3.4 shows an example of good progression. The time-space diagram (Figure 3.4a) shows perfectly ideal progression, with a series of platoons completely uninterrupted by intermittent red intervals. In the real world, such conditions are extremely rare, but similar conditions can occur when there is only one inbound flow to be dealt with. Figure 3.4b shows a PCD for an intersection at a diamond interchange where the inflow consists of through and right-turning vehicles, which form a platoon that is served almost entirely during green during the AM peak (5:00–9:00) and the PM peak (15:00–19:00). During other times of day, the platoons are more spread out and there are some arrivals in red. The difference is likely due to a different phase sequence at the upstream intersection.

- Figure 3.5 shows a case where there are no incoming platoons and arrivals are random. The time-space diagram (Figure 3.5a) shows that the vehicle trajectories are spaced out evenly, while the PCD (Figure 3.5b) shows that the dots are spread out fairly randomly within each cycle. This situation typically occurs at the first intersection in a corridor, when the closest nearby intersection is a few miles away or more. The first intersection has the role of forming the platoons that can be progressed through subsequent intersections.

- An example of poor progression is illustrated by Figure 3.6. In the time-space diagram (Figure 3.6a), many of the arrivals occur during red, meaning that the signal timing performs poorly. The PCD (Figure 3.6b) exhibits this behavior during the PM peak (15:00–19:00), as illustrated by the higher density of dots below green rather than above it. The other times of day have better arrival patterns, especially the AM peak (6:00–9:00).

- Oversaturation is characterized by long queues that cascade from one cycle to the next as volume exceeds capacity. This situation is illustrated in Figure 3.7a, where the queues do not clear during each cycle and the back of the queue seems to be growing longer as more and more vehicles stack up at the intersection. Figure 3.7b shows a corresponding PCD for a similar situation, which shows the detector occupancy in addition to the dots for the initial detection times. The arrival patterns have a distinctive pattern from around 16:00–17:30. The buildup of large queues over the detector is evidenced by the “empty” space between around 30 and 90 seconds into the cycle, where the queue has reached the advance detector. This alternates with a “busy” period during the rest of the cycle representing vehicles passing over the detector at saturation flow, as the queue discharges.
(a) Conceptual time-space diagram.

(b) Coordination diagram.

Figure 3.4. Traffic flow through an intersection for platoon vehicles with perfect progression.
Figure 3.5. Traffic flow through an intersection for randomly arriving vehicles.
Figure 3.6. Traffic flow through an intersection for platoon vehicles with poor progression.
(a) Conceptual time-space diagram.

(b) Coordination diagram with occupancy indicated by blue lines.

Figure 3.7. Traffic flow through an intersection for platoon vehicles with a capacity constraint.
3.2.2 Delays of Waiting Vehicles and Pedestrians

If the coordination diagrams presented earlier depict whether the agency has successfully attempted to “not stop traffic,” then a few additional metrics are needed to establish whether, if stopped, the agency has been successful at making traffic “not stop very long”. One way to do this is to measure the time between the first call for service on a movement, to the time that it takes to serve that movement [4,5].

Figure 3.8 illustrates the concept behind these measurements for vehicles (Figure 3.8a) and pedestrians (Figure 3.8b).

- For a vehicular movement, the actuation of the vehicle detector represents the arrival of a vehicle. After the start of green, the first time when the detector becomes unoccupied marks the latest moment when a vehicle might have left. Therefore, the time between the detector turning on during red and the detector turning off in green represents the maximum vehicle delay [5].
- For a pedestrian movement, the actuation of the pedestrian button represents the arrival of a pedestrian. After the start of the walk phase, any waiting pedestrians are assumed to proceed to traverse the intersection [4].

(a) Maximum vehicle delay.

(b) Pedestrian delay.

Figure 3.8. Delay estimate concepts.
Figure 3.9 shows a 24-hour calculation of the maximum vehicle delay. Superimposed on this plot are the red time and the cycle length. For this movement, the maximum vehicle delay is usually a little less than the red time for most of the day. Usually the waiting time lasts for 40–60 seconds. During the PM peak, around 15:00–19:00, the delay sometimes exceeds the red time. In seven cycles during the 24 hour period, the delay exceeds the cycle length, meaning that the detector has been continuously occupied for a very long time. This indicates high utilization of the phase at those times. The delay is particularly low during the overnight periods; during fully-actuated control, the controller serves the movement as quickly as possible and few other demands compete for service.

Figure 3.10 shows pedestrian delay over 24 hours. Because this shows the time between the pedestrian button actuation to the start of the walk phase during the successive green, the pedestrian delay is always less than the red time. Typical pedestrian waiting times are around 40 seconds, as long as 60 seconds during the PM peak hours (15:00–19:00). There are no pedestrians (or at least no one pushing the pedestrian button) in the early morning. This intersection operates in a fully-actuated mode for the entire day.

These metrics indicate the waiting time experienced by stopped vehicles and pedestrians at intersections. Incorporating detector data allows the utilization to be visually assessed (number of lines) as well as the magnitude of the delay (vertical height of the lines). This information would substantiate reports of excessive delay and facilitate evaluation of changes to signal timing plans, by showing whether delay increased or decreased in response to a revised control scheme.
Figure 3.9. Maximum vehicle delay.

Figure 3.10. Pedestrian delay.
3.3 **Corridor Level Analysis**

An analysis of arrival characteristics at the intersection level allows the analyst to investigate the local characteristics of system performance, such as whether the arrivals at one or more approaches coincide with the green times. A corridor contains many approaches at multiple intersections, each of which belongs to a multiplicity of possible routes in a network. The detailed analysis of operations at the intersection level can be coupled with higher level measurements of travel time along priority routes to evaluate corridor performance. This section discusses the relationship between arrival characteristics and corridor travel times and sensitivity of the measured corridor travel times to sensor locations.

3.3.1 **Relationship between Arrival Characteristics and Travel Times**

The relationship between vehicle arrivals and the state of the signal as vehicles arrive influences the quality of service experienced by those vehicles. Vehicles that arrive on red have to stop, while those that arrive on green do not (unless there is a queue present). As shown at the intersection level, arrival characteristics can be examined by relating arrivals to phase times using PCDs. At the corridor level, the collective effects of the arrival patterns at several intersections in succession on the travel time can be measured using vehicle re-identification.

Figure 3.11 shows a test network for investigating the relationship between arrival characteristics and travel time. This is a relatively small network of three signalized intersections along US 30 in northwest Indiana. The three signals are coordinated during most of the day (from 6:00 to midnight) and side-street demands vary strongly from one intersection to the next. Two Bluetooth monitoring devices (indicated by “BT Case”) were located on opposite ends of the corridor to measure vehicle travel times in both directions along the arterial route.

Figure 3.12 shows 24 hours of resulting travel times on Thursday, February 9, 2012 for the eastbound (Figure 3.12a) and westbound (Figure 3.12b) directions.

- In the eastbound direction, observed travel times are clustered into two distinct groups. There is one group with longer travel times (Figure 3.12a, callout “i”) and a second group with shorter travel times (Figure 3.12b, callout “ii”). There are more observations in the shorter travel time group for most of the day.
- The westbound direction sees a similar grouping of the observations into a longer travel time group (Figure 3.12b, callout “iii”) and a shorter travel time group (Figure 3.12b, callout “iv.”). The longer travel time group seems to have more observations during the afternoon. In general, there are more observations later in the day compared to the eastbound direction.
Figure 3.11. Travel time data collection on US 30 [6].
Figure 3.12. Raw travel times in the US 30 example, 2/9/2012 [6].
Figure 3.13 shows the corresponding PCDs for the three eastbound and three westbound approaches between the two travel time measurement points. In the eastbound direction, there are many arrivals at the first intersection (Figure 3.13a) during green, but the platoon extends into the red, as shown by the distribution of vehicle detections close to the horizontal axis. The duration of green is lower compared to the other eastbound PCDs. At the second intersection, the larger platoon arrives during green (Figure 3.13c) while a secondary platoon seems to arrive slightly earlier. Finally, at the third intersection (Figure 3.13e), a larger proportion of vehicles seems to make it through on green during most of the day, although there does still appear to be a fair number of vehicles in a secondary platoon that arrive during red.

Working through the westbound direction, most of the vehicles appear to arrive on green at both Int. 3 (Figure 3.13f) and at Int. 2 (Figure 3.13d). As for the eastbound direction, the westbound arrival pattern at Int. 1 (Figure 3.13b) has the most substantial number of arrivals on red, with a very substantial number of vehicle detections under the green line. This is especially true of the afternoon.

Compare the PCDs to the observed travel times shown in Figure 3.12. In both directions, there is a bimodal distribution of points, which corresponds to a portion of the traffic along the arterial being forced to stop while another portion does not have to stop. Analysis of the PCDs in Figure 3.13 shows that Int. 1 has a strong amount of arrivals on red in both directions, which is very likely to be the source of the stops and of the increased travel time. For example, the westbound movement at Int. 1 (Figure 3.13b) has a very wide swath of vehicle arrivals in red during the PM peak, which is also seen in the increase in travel times in Figure 3.12b.
Figure 3.13. Analysis of progression at Intersections along US 30, 2/9/2012 [6].
While these graphics demonstrate that the performance measures allow qualitative observations of the travel times, the real power of the methodology is shown when the performance measures are tracked over time.

Figure 3.14 shows two sets of cumulative frequency diagrams (CFDs) representing travel times during the AM peak (6:00–9:00) through four intersections on SR 37 in Noblesville, Indiana. Figure 3.14a shows northbound travel times while Figure 3.14b shows southbound travel times. Most of the CFDs are stacked on top of each other, with the exception of two highlighted CFDs that lie over to the right, meaning that travel times were longer on those days. Examination of the weather at the time found that there was snowfall on the day with the longest travel times, resulting in a 95 second increase in median travel times in the northbound direction, and 83 seconds in the southbound direction. The following day, travel times had recovered somewhat in both directions, but were still longer than the typical travel times.

During snow, vehicle speeds are reduced, so one would expect travel times to increase. However, the PCDs allow a more detailed investigation of why travel times increase. Figure 3.15 shows a map of the corridor and an array of PCDs for the southbound direction at the four intersections through which travel times were measured. The left PCD shows the snow day operation while the right PCD shows the clear weather operation. The PCDs are limited to the 6:00–9:00 time period to ease comparison. At the first intersection (I-01), arrivals are random in both cases, so there is no observable difference in the arrival patterns. The second intersection (I-02) sees the platoons arriving a little later in the cycle during the snow day, but most arrivals are still in green. The later arrivals correspond to speed reductions caused by snow on the pavement. At the third intersection (I-03), the arrivals are again shifted to later in the cycle, with the result that the ends of the platoons are cut off (callout “A”). This is not seen during clear conditions (callout “B”). Finally, at the fourth intersection (I-04), the arrival patterns are changed substantially, with arrivals mostly in red during the snow day (callout “C”) whereas during clear conditions they mostly occur in green (callout “D”).
Figure 3.14. Increase in travel time due to a winter weather event [7].
Figure 3.15. PCD comparison of AM Peak 06:00 – 09:00 during Thursday snow (7-Jan. 2010) and clear weather (14-Jan. 2010) conditions [7].
These examples demonstrate the utility of combining a high-level performance measure such as travel time with a more detailed performance measure such as the PCD. The CFDs of travel times are particularly useful for evaluating performance and quickly identifying when anomalies occur, while the PCD serves as a diagnostic tool for analyzing the performance of individual approaches in the corridor.

The analysis can be simplified by wrapping up the graphical performance measures into a single number. The PCD, for example, can be summarized with a single value for the overall percent on green (POG). The travel time CFDs, meanwhile, can be summarized by selecting the median value, and perhaps the 25th and 75th percentiles to describe the variability. This concept is explored further in the discussion of Network-level performance measures.

### 3.3.2 Travel Time Sensitivity to Sensor Location

The previous discussion showed corridor travel times and their relationship to individual approach performance. Because the locations of the sensors along the arterial affect the travel times that will be measured by the vehicle re-identification, a brief discussion is warranted. The following will examine results from a field study to facilitate comparison.

Figure 3.16 shows a map of SR 37 south of Indianapolis, and the location of several vehicle ID sensors (labeled as “BMS-1” through “BMS-8”) situated at midpoints between various intersections along the route. Three alternative methods of conducting matching between these sensing locations are displayed to the right. Origin-based matching computes the travel time from the entry point to various destinations, Destination-based matching computes the travel time from various entry points to the exit point, and the individual approach method computes travel times from one link to another, capturing travel time through each intersection.

Figure 3.17 compares a sample of PM peak hour travel times measured and tabulated by each method, with origin-based travel times in Figure 3.17a, destination-based travel times in Figure 3.17b, and individual approach travel times (converted to control delay) in Figure 3.17c. The origin-based and destination-based travel times increase as each sensor pair increases in distance. In Figure 3.17a, for example, the BMS-1 to BMS-2 data corresponds to the smallest segment under Figure 3.16a. In Figure 3.17b, BMS-1 to BMS-8 is the shortest span, under Figure 3.16b. Naturally, as distance increases, so does the travel time.

Note that as more signals are included in the route, the travel time tends to become less reliable (more variable). This is demonstrated by the increasing slope of the CFDs as they move further to the right in Figure 3.17a and Figure 3.17b. A slight “acceleration” of the slope increase can be seen between two particular CFDs in both plots. In Figure 3.17a, the slope of “BMS-1 to BMS-5” is more horizontal than “BMS-1 to BMS-4.” Meanwhile, in Figure 3.17b, the slope of “BMS-4 to BMS-8” is more horizontal than “BMS-5 to BMS-8.” These changes in slope are more rapid than between neighboring lines to the left or to the right, and they both correspond to the addition of the link between BMS-4 and BMS-5 in Figure 3.16, which traverses Southport Road.
Figure 3.16. Bluetooth monitoring stations and subsection segment regimes on SR 37 south [8].
Figure 3.17. Southbound SR 37 travel time analysis during the PM peak period (16:00–19:00) [8].
Figure 3.17c shows the control delays obtained by measuring travel times across the intersections. This is established by subtracting the 5th percentile free flow travel time from the distribution. Most of the CFDs of the delay show a median delay at a relatively low amount of about 15 seconds, but the segment “BMS-4 to BMS-5” has a much longer 75th percentile than the rest. This suggests that about a quarter of the measured vehicles on this route are experiencing an increase in delay, probably caused by stopping at Southport Road. From this, a possible conjecture is that Southport Road experiences some operational problems during the PM peak.

This is confirmed by digging into the high-resolution data from the same time period. Figure 3.18 shows three PCDs corresponding to the southbound movements at Banta Road (Figure 3.18a), Southport Road (Figure 3.18b), and Wicker Road (Figure 3.18c). The PCDs range over the 15:00–19:00 time period for a representative day. Of the three PCDs, Southport Road has the most distinct looking PCD, with practically every cycle having an arrival pattern that suggests that platoons are cut off by the start of red (Figure 3.18b). This contrasts with Wicker Road, where the platoon lies safely within the green time (Figure 3.18c) and with Banta Road, where most vehicles arrive on red but there does seem be a constraint (Figure 3.18a).

As would be expected, endpoint to endpoint travel time distributions can effectively characterize the overall corridor performance, but greater detail can be obtained by breaking long segments into shorter subsections. The greater the distance spanned by two endpoints, the smaller the impact of individual approach performance will appear. Given a long enough corridor, the impacts of signal timing changes will become almost imperceptible in the overall CFDs. This should be taken into consideration when considering where to locate sensors for field studies. In order to capture improvements, some midpoint travel time sensors should be included along with endpoints.
Figure 3.18. Southbound PCDs on SR 37 South during the travel time data collection.
3.4 **Network Level Analysis**

At the top level, the analyst is most interested in assessing overall system performance as a whole, and in searching across the inventory for problems to target resources. A network with 100 traffic signals, with 12 movements per intersection, and seven time-of-day plans covering a week has about 8,400 different movement/time-of-day combinations that could potentially be checked for deficiencies. This section demonstrates methods of sifting through this data to find where attention is most needed, and for aggregating across a large amount of data to produce overall system-wide metrics.

3.4.1 **Aggregations of Lower Level Data**

One approach to developing a system-wide picture of network performance is to aggregate granular pieces of data at the lower levels. If data is collected on an automated basis and continuously tabulated into performance measures, the task of assembling performance reports for multiple intersection perspectives is an incremental step.

For example, the intersection performance measure of the maximum vehicle delay presented in Figure 3.9 can be generated for each movement at each signalized intersection in the network where stop bar detection is used. By carrying out that tabulation over a given time period, the performance of each movement can be indexed and ranked at a system level to identify the most problematic system components. This can be repeated with other performance measures to focus on different objectives.

To facilitate an example, the maximum vehicle delay was calculated for 60 intersections in Indiana that were reporting complete data on May 31, 2014. A distribution of the median delays was assembled for each individual movement in the system, and the median value was selected as the metric to represent that movement in the system-level aggregation. The data was segmented by time of day to separate performance characteristics of peak hours and off-peak time periods.

Figure 3.19 shows the top 20 highest median maximum vehicle delay values by movement for the AM peak data segment (6:00–9:00). This type of ranked listing of system elements, independent of other attributes, is sometimes called “Pareto sorting.” The analysis locates one particular movement with a substantially higher value than the rest of the system during the same time period. The median delay value is over 100 seconds; the next movement has a delay of 86 seconds, and the rest begin to fall off gradually.

For comparison, the entire data set with 399 movements is shown in Figure 3.20, which provides a sense of the overall system performance during the AM peak. While this data view does not yet attach the performance to specific assets, it does provide some basic information, such as the proportion of movements with median delay values above, for example, 60 seconds. The same concept can be repeated with other performance measures to serve other objectives.
Figure 3.19. Median maximum vehicle delay by movement (highest 20 values).

Figure 3.20. Median maximum vehicle delay by movement (all movements).
Figure 3.21 presents the same data geographically. In this case, the data has been aggregated by intersection so that the intersection value represents the highest median delay value occurring at that intersection, in order to locate problems. Alternately, an average value or some other composite might be used, varying with objective. The intersection data is separated into five groups that allow locations with higher and lower delays to be distinguished. The map shows seven corridors around the greater Indianapolis area, and it is visually apparent that four corridors have most of their intersections in the highest group while the other three are in the lower group. There is also substantial variation within each corridor in some cases.
This aggregation example uses an intersection-level metric. However, the same approach can be taken to corridor-level data such as travel times. An example was developed in Indiana using estimated travel times based on crowd-sourced travel time data. As discussed earlier, this data set does not provide actual measured travel times, but having an estimation from a wide variety of corridors helps facilitate an example that could be implemented using other, higher fidelity data sets.

In this study, travel time estimates were established for a variety of arterial corridors throughout the state of Indiana based on an aggregation of minute-by-minute speeds provided by the data set. The result was a representative distribution of travel times for each corridor that could be indexed for comparison with others. The average value was selected as the representative value for the central tendency, while the standard deviation characterized the degree of variability (and hence the “unreliability” of the travel time). Both of these metrics were normalized by dividing by an ideal travel time, based on the speed limit.

Figure 3.22 shows a ranked list of the corridors involved in the study from smallest to greatest value of the normalized average travel time (Figure 3.22a) and the normalized standard deviation (Figure 3.22b). As with the intersection data, the ranking approach provides detailed information about the overall performance of the system and the location of the “neediest” system components.

It is possible to compare and contrast this information with the intersection data presented earlier:

- **SR 37 North** features a number of intersections in the upper two delay groups, according to Figure 3.21, and it also ranks close to the highest index values in both Figure 3.22a and Figure 3.22b. There are likely both opportunities to improve travel times as well as poorly performing individual movements at the intersections.
- **SR 37 South** has a similar number of high delay intersections in Figure 3.21, but in terms of travel time performance, it is close to the middle of the list in Figure 3.22a and Figure 3.22b. Thus, while there are some individual movements that could probably be assisted, the overall travel time performance does not stand out.
(a) Sorted by normalized average travel time.  
(b) Sorted by normalized standard deviation of travel time.

Figure 3.22. Arterial ranking: AM Peak (6:00–9:00).  
Data shown for all Wednesdays, January–July 2014. [9]
3.4.2 Analysis of Data at the Network Level

In addition to aggregations of data tabulated at lower levels, it is also possible to perform analysis of the raw data at the network level. One application of this would be to compare the performance of alternative routes through the network, by comparing vehicle re-identification data from multiple points in the system.

Figure 3.23 explains the procedure. This graph represents a simple network with four nodes and two routes under consideration that originate at node A and terminate at node D. One route passes through node B, and the other passes through node C. There are other possibilities not considered in the analysis.

- A vehicle is considered to use the node B route, if it is observed at nodes A, B, and D. The travel time on this route is measured by $t_D - t_A$, and is conditionally assigned to this route given a third observation of the same vehicle at node B, $t_B$.
- Similarly, a vehicle is considered to use the node C route, if observed at nodes A, C, and D. The travel time is still $t_D - t_A$ but the vehicle is assigned to this route if there is an observation at node C, $t_C$.
- Lastly, other vehicles observed at nodes A and D have travel times $t_D - t_A$, but they cannot be assigned to the node B or the node C route. While there is a chance that they did use one of those routes but were not observed, they also could have used some other alternative.

Figure 3.23. Conditional vehicle matching for determining route utilization.
This methodology was applied to the analysis of alternative routes during a major detour due to a road closure in Northwest Indiana. Figure 3.24 shows a map of the area affected by the closure. The official detour route is shown by the blue line, which is an alternative route using only state highways as signed by the state DOT. An “unofficial” detour route was also available using a mixture of state highways and local roads that took a more direct route. A “hybrid” route combining the two was also a likely routing candidate that incorporated most of the official detour but avoided the extraneous north loop by diverting from state highways to a local road. Finally, an alternative “local” route using mostly local streets (the east-west road a decommissioned state highway) was a possibility. The map indicates the locations of the vehicle identifying sensors throughout the network.

Figure 3.24. Example network data collection for route analysis in Northwest Indiana [10].
The results of the analysis are shown in Figure 3.25. These are box-whisker plots of the travel times experienced on each of the four alternative routes for eastbound (Figure 3.25a) and westbound (Figure 3.25b) commuters. The data is further segmented into weekday and weekend groups for comparison. In addition to the travel times, the proportion of observations is also given for each of the four routes.

The results indicate that the official route was the least popular route used by commuters, with approximately 1 in 10 or fewer observed vehicles making use of this route. The travel times were also generally higher on this route than for the others, especially for westbound travel, which is not unexpected since this route covers a longer distance than the others. The local route ended up being the most popular route for making the detour.

For comparison, the travel time on each route available obtained from Google Maps at the same time as the study is also included. While in some cases it is similar, in others it is substantially different, especially for westbound movement on the official route. This points to the difficulty in assessing travel times on arterial networks through estimation. Although the travel time metrics used by applications for the general public may have improved since the time of the study, the divergence from the field measurement highlights the need for agencies to make independent measurements.
(a) Eastbound route.

(b) Westbound route.

Figure 3.25. Travel times for alternative routes [10].
3.5 Guidance for Operational Analysis

To conclude this module, a sampling of case studies are presented where vehicle re-identification data was combined with high resolution signal event data to achieve a performance objective for an arterial system. The first case study considers the optimization of progression along an arterial corridor, and the second case study tracks the longitudinal performance of a system to determine whether the signal timing plan is aging. Finally, results from the second case study are further applied to quantify user benefits resulting from operational improvements.

3.5.1 Case Study: Signal Timing Optimization over the Weekend

A common objective of arterial operations is to coordinate intersections to promote smooth traffic flow, increasing the chances that a vehicle would be able to pass from one end of the corridor to the other without stopping. Often, different timing plans are devised for weekday and weekend conditions. However, unless the traffic engineering staff has dedicated weekend time to observe weekend operations and make changes, weekend conditions will generally be unknown. The same thing might be said for certain times of day as well.

This case study [3] examines a timing plan that operates on Saturdays, identifies an operational deficiency related to the objective of providing good progression, and validates an improvement, all of which could be achieved remotely and without having to invest time observing conditions in person during the timing plan.

Figure 3.26 shows a map of the system. Four intersections are shown, although at the time of the study only the four intersections on the north end of the system were equipped to log the high-resolution signal data. Vehicle sensors were located at the two endpoints of the system as indicated.
Figure 3.26. Map of a four-intersection span of SR 37 in Noblesville, Indiana [3].
The initial, raw travel time data is shown in Figure 3.27. The macroscopic view on the data shows that the Saturday travel times are higher than any other day of the week. The point cloud within the Saturdays tend to be about 2 minutes higher than the weekdays, or Sundays. This suggests that an operational deficiency likely exists and that Saturdays represent a candidate time period for further investigation. This demonstrates the utility of automated travel time measurements for making a high level analysis possible. It also demonstrates that the data itself does not necessarily show the exact cause of deficiencies, but it does enable the analyst to quickly prioritize the time periods where such problems exist. It would be difficult to achieve the same type of information by manual observations, and there were no complaint calls that would have initiated a field investigation.

Figure 3.27. Northbound Travel times along SR 37 prior to offset adjustment [3].
To investigate the conditions in more detail, high-resolution data was collected from the four intersections on the north end of the system and PCDs were constructed to determine what was occurring on Saturdays. Figure 3.28 shows the resulting eight PCDs from a representative Saturday. A number of observations can be made from these:

- A good quality of progression exists on the southbound approach at Town and Country Blvd. (Figure 3.28c) and the northbound approaches at SR 32 (Figure 3.28e) and Town and Country Blvd. (Figure 3.28g) as demonstrated by most of the vehicles arriving in green (the point cloud is mostly above the green line).
- The quality of progression is poor on the northbound approaches at Pleasant St. (Figure 3.28f) and Greenfield Ave. (Figure 3.28h), as shown by most of the vehicles arriving in red (the point cloud is mostly below the green line). The southbound approach at Greenfield Ave. (Figure 3.28d) shows arrivals just prior to green. A better situation would have those vehicles arrive after the start of green. The southbound approach at Pleasant St. (Figure 3.28b) shows a similar situation, although the arrival distribution appears more spread out because of a heavy side street entry.
- There are random arrivals on the southbound approach at SR 32 (Figure 3.28a). This is not unexpected since this is the entry point into the system.
- Arrivals are not random on the northbound approach at Greenfield Ave. (Figure 3.28h), even though it is about 1.5 miles from the next intersection. This defies the common rule of thumb that one mile is about the upper limit to coordinating neighboring intersections, because there are clearly well formed platoons arriving at the end of this link which would benefit from coordinating well. The reason for this is that the upstream link has no driveways, and with the lack of “friction” from driveway entries and exits, combined with a speed limit of 55 mph, platoons are well-preserved over a relatively long distance.

The analysis shows that there are opportunities to potentially improve progression on three of the approaches in the system. The poor conditions at two northbound approaches are very likely the cause of the increased travel time for the northbound direction in the corridor.
Figure 3.28. Coordination diagrams prior to adjustment [3].
Based on this analysis, the Saturday timing plan was updated to improve the offsets in order to improve traffic flow through the corridor. In this study, the adjustments to the offsets were determined using a simple prediction of how the vehicle arrivals would change in response (the details of which are provided elsewhere). However, the signal timing could have been improved by any number of methods according to agency policy. More importantly, the previously unknown deficiency was identified and prioritized for action without requiring personnel to make field observations during the Saturday timing plan.

Figure 3.29 shows the outcome of the retiming effort.
- The two northbound approaches at Pleasant St. (Figure 3.29f) and Greenfield Ave. (Figure 3.29h) now have most of their arrivals occurring during green rather than in red, which is a total reversal from the previous conditions.
- The southbound approaches at Pleasant St. (Figure 3.29b) and Greenfield Ave. (Figure 3.29d) have been slightly adjusted so that the bulk of the platoons now coincide with the green, whereas they arrived slightly prior to green previously.
- The northbound approach at SR 32 (Figure 3.29e) now sees arrivals slightly prior to green, which is a little worse than the previous condition. The tradeoff is in the slight improvement for the southbound vehicles at Pleasant St. (Figure 3.29b). When making adjustments to offsets, there is often a tradeoff between opposing movements along a link similar to this one, and the analyst has to make a decision regarding which movement to prioritize.

The impact on corridor travel times is shown by a view of the raw data in Figure 3.30, and by comparison of CFDs in Figure 3.31.
- Figure 3.30 shows that, during the weeks prior to the offset adjustment, the point cloud for Saturday travel times are now comparable with the other days of the week.
- Figure 3.31 represents the travel time data with CFDs that are stacked on top of each other to facilitate comparison. Eliminating the two stops for movement in the northbound direction led to a 1.9-minute improvement in the median northbound travel time (Figure 3.31a). At the same time, there was a more modest reduction in the median southbound travel times of about half a minute (Figure 3.31b).

Thus, by bringing together travel time data with high resolution data, it was possible to target engineering resources to improve an operational deficiency, and to quantify the degree of benefit for the traveling public.
(a) Southbound at SR 32.
(b) Southbound at Pleasant St.
(c) Southbound at Town and Country Blvd.
(d) Southbound at Greenfield Ave.
(e) Northbound at SR 32.
(f) Northbound at Pleasant St.
(g) Northbound at Town and Country Blvd.
(h) Northbound at Greenfield Ave.

Figure 3.29. Coordination diagrams after adjustment [3].
Figure 3.30. Northbound Travel times along SR 37 after offset adjustment [3].
Figure 3.31. Saturday travel time cumulative distribution functions (from 0900 to 1300) [3].
3.5.2 Longitudinal Tracking to Investigate Timing Plan Aging

In addition to before-after studies, tracking performance over time enables the analyst to determine whether conditions are changing over time and, if so, the rate of the change. This is more powerful than a limited before-after study because it determines whether one-time changes have a lasting effect and how quickly the control plan is aging. This section presents a few examples facilitated by travel time data and high-resolution signal event data.

Figure 3.32 shows a map of US 31 through Kokomo, Indiana. This roadway serves as the major north-south arterial in the city of Kokomo. At the time of the study [11], it was also the only route through the city (a bypass was constructed a few years later). There are 13 intersections along this corridor. In response to known shortcomings of the existing timing plan, the responsible agency made a concentrated effort to develop a new timing plan, which was implemented and accompanied by several days of post-implementation tuning.

The outcome of the retiming effort is illustrated by a collection of “before” and “after” travel time CFDs in Figure 3.33. These illustrate conditions prior and subsequent to the retiming project. These are separated according to the five different time of day plans in use along the corridor. Notably, within every single segment of the data, travel times were improved by a substantial amount, over 1 minute in most cases. In contrast to the previous CFD comparisons presented in this module, the CFDs here represent not only the “average” before/after conditions but also the month-to-month performance across an extended time period prior to and after the improvement.

Figure 3.34 shows the median values of the travel time as they changed, before and after the retiming activity. The data here consists of an aggregated travel time estimate constructed using all of the data available within each month. Similar conditions are observed for both the southbound (Figure 3.34a) and northbound (Figure 3.34b) directions of travel. After the retiming efforts, which occurred in April 2012, travel times were substantially reduced in May 2012. However, in the next few months that followed, the margin of improvement was slightly reduced as travel times began to rise again. This proved to be a seasonal condition, because the travel times again fell during the following fall and winter.

Overall, the results show that there was no significant aging in the timing plan for a year after its implementation, and additionally that the trends were similar for both directions of travel on the corridor.
Figure 3.32. US 31 in Kokomo, Indiana [11].
Figure 3.33. Monthly travel time CFDs from before and after retiming [11].
Figure 3.34. Tracking monthly travel time CFDs over several months after retiming [11].
A similar longitudinal analysis can be conducted using the high-resolution data. As discussed earlier, the percent on green characterizes the overall quality of progression on an approach as a single value that aggregates all of the data represented in a coordination diagram in disaggregate terms.

High resolution data has been collected over the span of several years at several locations in Indiana. Given a large library of historical data to work from, it is possible to calculate changes to the performance over time which would determine whether timing plans are continuing to operate well or whether conditions are growing worse. Here, the percent on green (POG) is shown for two different signal approaches over a two-year span.

Figure 3.35 shows how POG changed over time from March 2013 through February 2015 at a new intersection that was added to an existing arterial. The signal was activated in early 2013. Figure 3.35a shows the performance of the northbound approach, while Figure 3.35b looks at the southbound approach. For this comparison, data from an off-peak timing plan (0900-1100) is examined. Early in the operation, the POG is close to 100% for several months in the northbound direction (Figure 3.35a), but gradually starts decreasing over time to a value of 70%. For the southbound direction, the POG is above 90% for almost all of 2013 and 2014, but starting in late 2014 and the first few months of 2015, it starts to fall below 90%.

The POG was initially very high because this was a brand new intersection. While there were some existing businesses served by the intersection, these did not generate much initial traffic. However, gradually, new facilities were added nearby, and drivers began to increase their usage of the intersection. Thus, early on, the lack of side street activity meant that the signal would tend to dwell in green on the mainline, leading to a high POG. As side street activity has increased, more and more green time was transferred from the mainline to the side street, thus lowering POG.

The gradual degradation of the POG indicates that this intersection should be considered for signal timing adjustments to accommodate the changing traffic patterns. The established signal timing, based on field tuning under low traffic volume conditions, may not adequately serve traffic under changing conditions. Module 4 will explore this concept in additional detail while investigating the effectiveness of capacity allocation throughout an arterial network.
Figure 3.35. Tracking intersection performance over time.


3.5.3 Quantifying User Benefits

Engineers are often asked to provide an estimate of the benefits obtained from investments of resources into system operation. Thus, this module closes with an example of how automated measurements of system performance can facilitate such a comparison. Here, we look again to the Kokomo retiming project discussed in the previous section. The median travel time values from before and after periods were tabulated to come up with a change for all of the timing plans in use on site. This included not only the five coordinated time of day plans examined previously, but also the overnight periods.

The median travel time changes were multiplied by the volume of traffic making use of the movement to derive a value of the total savings. The ADT of the corridor was 26,000 per day, and this was segmented into representative numbers by time of day using a proportion of ADT. The daily value obtained from these numbers was then annualized by multiplying by 260 weekdays occurring over a 52-week period. These were converted into monetary units using the methodology used in the TTI Mobility Report [12], using the following formulas [13].

First, the number of hours \((h)\) was found from

\[
 h = \left( f_{ADT} \right) \left( ADT \right) \left( \frac{\Delta TT_{min}}{60} \right) (D),
\]

\text{Equation 3}

where \(f_{ADT}\) is the fraction of ADT associated with the time segment (by direction and time of day), \(ADT\) is the ADT (number of vehicles), \(\Delta TT_{min}\) is the change in travel time (in minutes), the factor of 60 converts this to hours, and \(D\) is the number of days to be considered. Here, because weekdays were used, \(D = 260\).

Next, this was converted into a user benefit \((U)\) by

\[
 U = \left( h \right) \left( PC \right) \left( x_{car} \right) \left( p_{car} \right) + \left( h \right) \left( 1 - PC \right) \left( x_{truck} \right),
\]

\text{Equation 4}

where \(PC\) represents the percentage of cars, \(x_{car}\) and \(x_{truck}\) are the monetary value of time (in dollars per hour) for passenger cars and trucks respectively, and \(p_{car}\) is the number of passengers per car. Values of $16.89, $86.81, and 1.25 were respectively assumed for \(x_{car}\), \(x_{truck}\), and \(p_{car}\).

The results are summarized in Table 3.1. Because there was a reduction of the travel time for both directions for every time of day, a positive result was obtained in each case. The total annualized travel time benefit was equal to approximately $2.7 million.
Table 3.1. Annualized savings due to signal retiming [11].

<table>
<thead>
<tr>
<th></th>
<th>Plan Median TT Savings (min.)</th>
<th>% of Daily Traffic</th>
<th>TT Savings (hrs.)</th>
<th>TTI Travel Time Savings ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Southbound US 31</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plan 0 (0000 – 0500)</td>
<td>0.79</td>
<td>2.2%</td>
<td>1987.34</td>
<td>$ 46,941.69</td>
</tr>
<tr>
<td>Plan 1 (0500 – 0900)</td>
<td>1.22</td>
<td>7.2%</td>
<td>9925.88</td>
<td>$234,453.24</td>
</tr>
<tr>
<td>Plan 2 (0900 – 1100)</td>
<td>1.83</td>
<td>5.3%</td>
<td>10877.93</td>
<td>$256,941.12</td>
</tr>
<tr>
<td>Plan 3 (1100 – 1300)</td>
<td>1.1</td>
<td>6.7%</td>
<td>8246.25</td>
<td>$194,779.77</td>
</tr>
<tr>
<td>Plan 4 (1300 – 1500)</td>
<td>0.93</td>
<td>6.6%</td>
<td>6886.14</td>
<td>$162,653.47</td>
</tr>
<tr>
<td>Plan 5 (1500 – 1900)</td>
<td>1.53</td>
<td>13.5%</td>
<td>23311.22</td>
<td>$550,620.34</td>
</tr>
<tr>
<td>Plan 6 (1900 – 2400)</td>
<td>0.91</td>
<td>7.1%</td>
<td>7319.89</td>
<td>$172,898.62</td>
</tr>
<tr>
<td><strong>Northbound US 31</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plan 0 (0000 – 0500)</td>
<td>0.58</td>
<td>2.2%</td>
<td>1462.30</td>
<td>$34,540.02</td>
</tr>
<tr>
<td>Plan 1 (0500 – 0900)</td>
<td>0.75</td>
<td>7.6%</td>
<td>6420.27</td>
<td>$151,649.25</td>
</tr>
<tr>
<td>Plan 2 (0900 – 1100)</td>
<td>1.02</td>
<td>5.5%</td>
<td>6316.57</td>
<td>$149,199.92</td>
</tr>
<tr>
<td>Plan 3 (1100 – 1300)</td>
<td>1.1</td>
<td>7.0%</td>
<td>8627.08</td>
<td>$203,775.18</td>
</tr>
<tr>
<td>Plan 4 (1300 – 1500)</td>
<td>1.26</td>
<td>7.0%</td>
<td>9881.93</td>
<td>$233,415.21</td>
</tr>
<tr>
<td>Plan 5 (1500 – 1900)</td>
<td>0.69</td>
<td>14.2%</td>
<td>11040.76</td>
<td>$260,787.26</td>
</tr>
<tr>
<td>Plan 6 (1900 – 2400)</td>
<td>0.45</td>
<td>7.9%</td>
<td>4018.01</td>
<td>$94,906.91</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td>100.0%</td>
<td>116321.6</td>
</tr>
</tbody>
</table>
3.6 **Summary**

This module examined the application of vehicle re-identification data and high-resolution traffic event data towards achieving objectives related to travel time and travel time reliability. Observable elements of system performance were examined at the intersection, corridor, and network levels.

- At the intersection level, a useful performance metric for examining the quality of progression is the Purdue Coordination Diagram (PCD), which graphically displays the details of detected vehicle arrivals relative to the actual green times that are served. This enables a quick visual check to determine whether vehicles are mostly arriving on green or on red. This can be tied with a single quantitative number, the percent on green (POG). Additional performance measures include estimates of delay such as the maximum vehicle delay for vehicle movements with stop bar detection, and the actuation to service time for pedestrian movements.

- At the corridor level, travel times in both directions along a particular route capture the performance of signal coordination and the quality of service for commuters who use the route. There is some value to representing travel times as a point cloud of observations, but a more succinct representation is the cumulative frequency diagram (CFD) of the observed travel times, which enables different time periods to be overlaid on top of each other for comparison purposes and identification of outliers.

- At the network level, there are opportunities to view data as an aggregation of the intersection and corridor level metrics. This approach enables “hot spots” of deficient performance to be located across the entire network quickly. Additionally, the raw data can be re-applied to a network level analysis to obtain additional details such as origin-destination properties.

The module concluded by presenting case studies where the data was applied in operations for objectives of achieving high quality signal progression. In the first case study, travel time data based on vehicle re-identification was used to identify a deficiency. High resolution data was then used to identify the necessary details needed to identify the causes of the deficiency and resolve them. After adjustments were made to the operation, the high resolution data verified the impact of the changes and the travel time data indicated the improvement to system performance. In the second case study, travel time data was compiled over a long time period to observe not only changes before and after a signal re-timing project, but also to check whether the improvements held up over time. Finally, the travel time data made possible an estimate of user benefits from retiming.
3.7 References

Module 4. Intersection Capacity and Demand

4.1 Introduction

This module examines the measurement of capacity allocation and utilization at intersections in an arterial corridor, from the performance of individual intersections and at the corridor and network level scales. The objective is to evaluate the efficiency of the local control at signalized intersections, and to determine whether the effects on the network as a whole. While Module 3 focused largely on the topic of progression, this module focuses more on identifying operational deficiencies where the demand exceeds the provided capacity.

In addition to identifying capacity deficiencies on a wide scale, the performance measures are also used to differentiate “correctable” deficiencies, which can be improved by re-allocation of green time (through signal timing adjustments), by identifying where spare capacity exists at intersections. This process also identifies deficiencies that are not correctable, which would require capital improvements (such as addition of new lanes) to address.

4.2 Intersection Level

At the core of intersection operations is the need to divide time between competing movements. For signal control, the “eight-phase” scheme is in widespread use, such that it can serve as a template for organizing the performance of the competing movements. Figure 4.1 shows an example intersection operated under eight-phase control, which is used to present example performance measures in this section.

In the eight-phase scheme, the controller operates both rings simultaneously, and phases in different rings are compatible with each other if they are on the same side of the “barriers” (thick black lines). For example, phase 1 can run at the same time as phases 5 and 6. In practice, the phase order may vary by time of day, and may dynamically change under some control schemes. However, the eight-phase scheme provides a useful layout for intersection performance measures to succinctly summarize the eight conventional competing demands, as will be demonstrated in this section.
Figure 4.1. An example eight-phase intersection (SR 37 and SR 32 in Noblesville, Indiana) [1].
4.2.1 Evidence of Oversaturation and Undersaturation in Microscopic Data

As mentioned at the start of Module 3, there is some interplay between the quality of signal coordination and the relationship between demand and capacity. Module 3 introduced the Purdue Coordination Diagram (PCD) as a means of visualizing vehicle arrivals relative to the green times on a signal approach. In addition to identifying this relationship, the PCD also shows conditions of oversaturation and undersaturation.

- Figure 4.2a shows 24 hours of operation at a signalized approach that receives traffic exiting from a freeway ramp. Throughout most of the day, the arrivals are randomly distributed throughout the cycle, which is expected given that there is no upstream signal control to packet the traffic into platoons. However, during the afternoon time period between 15:00 and about 18:00, the vehicle detections take on a repeating pattern. This occurs when queues at the intersection reach the detector, so the detection profile no longer represents arrivals, but rather the times when queue discharge reaches the detector. The “empty” region around the green line (Figure 4.2a, callout “i”) represents the time when queues are standing on top of the detector, and the densely packed vehicle detections at other times show the saturation flow discharge of queues during green. The buildup of a queue is a symptom of oversaturation.

- Figure 4.2b shows 24 hours of operation at a signalized approach in a much less congested area. In fact, the approach is only coordinated during the 15:00–19:00 time period; the rest of the day, it runs in a free (fully-actuated, non-coordinated) mode. The PCD shows that coordination is successful, with most of the arrivals occurring during green (i.e., most of the points are above the green line). However, there is a considerable portion of the green time where very few vehicles are actually arriving (Figure 4.2b, callout “ii”). This underutilized green time could possibly be used to serve another phase.

4.2.2 Quantifying Capacity

To investigate the existence of oversaturation and undersaturation at intersections more systematically, quantitative measurements of the capacity and the demand can be made. Capacity relates to the green time that is given to each movement at the intersection. The actual green time is controlled by the detector actuation logic and the coordination logic. During coordinated operations, each phase is given a share of the cycle length called the “split” that, if each phase is served the maximum amount of green in a cycle, would represent the combined amount of green time and clearance time given to the phase. However, because of actuation, some phases finish service early, allowing them to give up time that can potentially be used by other phases, depending on the controller configuration.

Figure 4.3 illustrates variations in capacity during one cycle. For simplicity, values are displayed for the upper ring (phases 1,2,3,4) only. The splits, as programmed in the controller, are displayed in Figure 4.3a as a percentage of the cycle, and in Figure 4.3b in the total number of seconds. The split time includes the green, yellow clearance, and all-red clearance time associated with a phase. Figure 4.3c shows the estimated green time based on the splits, with all of the yellow and all-red times pooled into the category of “lost” time. For example, based on the programmed splits, phase 2 would be expected to receive 32.5 seconds of actual green time during a given cycle operating under this coordination plan.
(a) SR 37 and 126th St., Northbound, 2/4/2015.

(b) SR 37 and 191st St., Northbound, 2/4/2015.

Figure 4.2. Examples of oversaturation and underutilization.
Figure 4.3. A comparison of estimated capacity to observed capacity for ring 1 (phases 1–4) for one cycle [2].
Based on the estimated green, the estimated capacity per phase is given by Figure 4.3d, assuming a saturation flow rate of 1900 vehicles per hour per lane (veh/h/ln). The formula for what might be called the equivalent hourly capacity \(c\) is

\[ c = s \cdot \frac{g}{C}, \]  

Equation 5

where \(s\) is the saturation flow rate, \(g\) is the green time, and \(C\) is the cycle length. Using Figure 4.3c as an example, given \(g = 32.5\), \(C = 96\), and \(s = 1900\) veh/h/ln, the equivalent hourly capacity is 644 veh/h/ln.

Outside of dense urban areas, few traffic signals operate on a fixed-time basis with each phase receiving the same green time in each cycle. Actuation adjusts the green times according to the traffic demand. The impact of actuation is illustrated in Figure 4.3e, which shows the actual green times during a given cycle. Note that phase 3 has disappeared from this view; during this cycle, there was no call for service for this phase, so it was skipped. The lost time is only 10.8 seconds rather than the 14.8 seconds in Figure 4.3c, because phase 3 was not served. The actual green times of phase 1 and phase 4 are lower than the estimated greens in Figure 4.3c, because those phases terminated early because of actuation. Phase 2 absorbed the time given up, swelling its green time to 53.4 seconds. The resulting “observed” capacities during this cycle are shown in Figure 4.3f, which applies the same assumptions of saturation flow to the actual greens rather than the estimated greens.

Because the high-resolution traffic event data includes a record of timestamps marking when each phase changes its state, the green times can be measured directly from the data, as shown in Figure 4.4. Each phase state change is marked by an event, and by comparing timestamps for the end of green and start of green, the duration of the green interval can be calculated. It is also possible to differentiate between the minimum green and the extended green, and so forth.

Some example data showing green time and capacity for a single phase are presented in Figure 4.5. The data represents Phase 6 at SR 32 and SR 37 (Figure 4.1). The amount of green time served in each cycle is shown in Figure 4.5a, while Figure 4.5b shows the corresponding amount of allocated capacity. The data represents a coordinated phase serving two lanes of a through movement. There is a considerable amount of variation in the green time because of actuation. The impacts of minimum times can be seen in the somewhat stratified distribution of points during certain times of day.
Figure 4.4. Measuring capacity.
Figure 4.5. Green time and capacity of a single phase.

(a) Green time.

(b) Equivalent hourly capacity.
Figure 4.6 expands the analysis to eight phases at the intersection, with eight plots of the capacity over 24 hours. The layout of the charts is inspired by the ring diagram concept (Figure 4.1). This representation of the data enables visualization of the relative emphasis of each phase at the intersection in terms of the provided capacity. For example, phases 1 and 7 stand out with relatively lower amounts of actual capacity than some of the others. Phases 2 and 6, which are the coordinated phases, stand out as having the highest amounts of capacity.

4.2.3 Quantifying Demand

Capacity only tells part of the story about intersection operations. The utilization of the capacity is the other part. With vehicle detection, it is possible to measure capacity utilization along with its allocation by means of the high-resolution data. Depending on the type of detection available, the sort of data available to the analyst varies.

- Where count data is available, it is possible to calculate volumes directly. Count detectors may exist at the stop bar. Advance (setback) detector data and other detection zones with footprints smaller than a vehicle length can also be used for count data.
- Where presence data is available from a stop bar detection, the occupancy can be used to estimate utilization.

Figure 4.7 shows the equivalent hourly volumes for eight phases at the same intersection for which the provided capacities were presented earlier. This visualization of the demands enables a visual comparison of the strength of demand from various components. The coordinated phases, 2 and 6, have very strong patterns of peaks during the AM and PM periods. The southbound movement dominates in the morning while the northbound movement is heavier in the afternoon. There is a similar but less prominent trend in phases 4 and 8, which represent eastbound and westbound movements. Demand for the left turns are mostly evenly distributed throughout the day. Phases 3 and 5 are relatively strong movements while 1 and 7 are weaker.

Where counts are available, it is possible to compute a volume-to-capacity (v/c) ratio from the following formula.

\[
\frac{v}{c} = \frac{N}{g \cdot s_s}
\]

Equation 6

Here, \(N\) is the number of vehicles, \(g\) is the green time, and \(s_s\) is the saturation flow rate in vehicles per second. The count, \(N\), takes place during both the green and the preceding red in order to include all of the vehicles that were served during the green. For an advance detector, the vehicle counts are adjusted for travel time from the detector to the intersection.
Figure 4.6. Capacities by phase assuming a 1900 veh/h/lane saturation flow rate.

Figure 4.7. Volumes by phase.
The volumes and the provided capacities are combined together in Figure 4.8, which shows the v/c ratio for the eight phases. The red line for $v/c = 1.0$ shows where the volumes exceed the theoretical capacity, which indicates oversaturation (split failures). A quick visual comparison of the phases reveals that phases 3, 4, and 5 seem to have the highest degree of utilization, while phase 8 is the lowest. Phase 1 and 7 have occasional cycles where the v/c ratio is high, but most of the time their utilization is lower. Phases 2 and 6 reach capacity during a few peaks but otherwise do not seem to have serious capacity deficiencies.

The same data is presented using a CFD of the cycle-by-cycle values in Figure 4.9. The plot for each phase shows one CFD tabulated for 24 hours of operation, and another that focuses only on the 6 am – 6 pm time period. The amount of oversaturation on each phase is indicated by the amount of the curve that extends beyond the $v/c = 1.0$ line. For example, phase 5 crosses this line at around the 80th percentile or so, which means that about 1 in 5 cycles has a higher demand than the provided volume.
Figure 4.8. Volume-to-Capacity ratio by phase.

Figure 4.9. CFD of v/c ratios over 24 hours by phase.
An alternative approach uses stop bar occupancy rather than counts. This may be a better approach for intersections where count data cannot be obtained using the detection system or where its accuracy is doubtful. Instead, this metric uses the same type of information used by the controller as it decides whether or not to terminate a phase. When a vehicle is present in the detection zone (that is, when the detector is occupied), the controller will continue to extend the phase until reaching the maximum time. Similarly, the measurement of the occupancy can be used to evaluate the amount of green that coincides with a traffic demand. Because all of the detector activation and termination times are recorded in the high resolution data, these occupancy values can be calculated with a high degree of precision.

Figure 4.10 explains the concept behind a metric based on occupancy [3]. The graphic in Figure 4.10a shows a timeline of events including two different detectors representing two through lanes at a signalized approach and their “on” and “off” states over time. This is followed by a chart of their combined “on” state (that is, if either detector is on, then the two detectors are collectively considered to be “on”). The phase state is indicated at the bottom; just above this, the information from the phase data and the detector data is combined to calculate the occupancy “ratios”, or the proportion of time during a given interval where the detector is occupied. The green occupancy ratio (GOR) is the amount of green with occupancy, and the red occupancy ratio (ROR) is the amount of red – in this case limited to the first five seconds – with occupancy.

The subsequent series of images show conditions in the field at different times. At the start of green (Figure 4.10b), there is clearly a very long queue waiting for service. One vehicle is marked by callout “V” that is near, but not quite at the end of the queue. The end of green comes nearly 30 seconds later (Figure 4.10c), and vehicle “V” has advanced almost to the stop bar, but not quite. At the end of the yellow clearance interval (Figure 4.10d), this vehicle has not quite made it and the driver has decided to stop, as confirmed by its presence in the detection zone five seconds after the start of red (Figure 4.10e). This is an example of a split failure, where the amount of demand during the cycle is greater than the provided capacity.

The combined ROR/GOR metric indicates both high utilization of the green (high GOR) with presence of vehicles immediately after the end of phase service (high ROR). Figure 4.11 shows a plot of ROR versus GOR for one phase over 24 hours of operation. Those cycles in which both ROR and GOR had high values (above 0.8) represent likely split failures. Thus, the more cycles for which data is clustered in the upper right hand corner, the more capacity deficiencies experienced by the phase.
Figure 4.10. GOR and ROR calculation for a single cycle of an oversaturated movement [3].
Figure 4.11. Plot of Red Occupancy Ratio (ROR) versus Green Occupancy Ratio (GOR).
Figure 4.12 expands the view of ROR/GOR across eight phases at the same intersection. As before, the clustering of points in the upper right hand corner indicates a prevalence of oversaturation on the phase. Figure 4.13 simplifies the view and attaches the points to a particular time of day by plotting the number of cycles per 30 minutes for which ROR and GOR both exceed 0.80. This represents the number of likely split failures. The data represents the same 24 hours of operation as in the volume-to-capacity metrics in Figure 4.8.

For the most part the same observations can be made based on the raw data alone. Phases 3, 4, and 5 have the most split failures, and these occur throughout the day. Phases 2 and 6 experience some split failures which are more prevalent during peak periods. Finally, phases 1, 7, and 8 appear not to have a significant number of split failures. There are none at all for phase 8, while phase 7 has only a handful, and phase 1 has a modest number during the AM peak but otherwise very few.

These graphics make it possible to visually distinguish intersection performance throughout the day very rapidly. It is not difficult to imagine the engineer responding to a complaint about inadequate green time on a particular movement to use this data to validate whether such a condition was likely in the field at the time of the complaint. Additionally, continuous and automated data processing would make it possible to instantaneously pull records from different dates to see whether conditions of oversaturation were recurrent or not. Finally, reports based on aggregations of this data might enable the engineer to proactively address emerging manifestations of oversaturation rather than waiting for complaint calls to be generated.

4.2.4 Capacity Utilization across the Intersection

Combining information about capacity allocation and its utilization is helpful for identifying hot spots of congestion at an intersection. However, it is perhaps more helpful to understand whether or not the problems can be adequately addressed by signal timing changes or whether more substantial investments in resources are needed.

A typical approach to eliminating split failures on a phase would be to look for another phase that could give up green time (that is, where the utilization is low), and transfer some of the cycle time from one split to another. A performance measure representing the overall intersection utilization would allow the analyst to determine whether spare capacity exists. The degree of intersection utilization, \( X_c \), is a composite metric that is based on volume alone (independently of the green times) which gives a sense of the total intersection utilization with respect to the total possible capacity allocation within a cycle. Details of the calculation are provided elsewhere [1]. Essentially, a high value of \( X_c \) indicates that the intersection is becoming completely oversaturated, in that the total possible capacity is inadequate. A lower value of \( X_c \) indicates that some spare capacity exists. When oversaturation coincides with a low \( X_c \) value, this means that there are some split failures that could potentially be corrected by transferring green time between phases.
Figure 4.12. ROR/GOR by phase.

Figure 4.13. Number of split failures per half hour, by phase.

Arterial Performance Measures Report
Contract #: DTFH61-14-C-00035

148
Figure 4.14 shows the degree of intersection saturation over 24 hours at the example intersection. Not surprisingly, the metric reaches its high points during the AM and PM peaks, when overall volumes are higher than the rest of the day. There are a handful of cycles where total volume exceeds the total potential capacity. During other time periods, the level of utilization hovers between about 0.50 and 0.75.

The higher the value of intersection utilization, the lower the likelihood that there is spare capacity that could be re-allocated to eliminate split failures. Thus, a “threshold” value can be selected to determine the time periods where spare capacity is most likely available and where there is strong possibility for a good return on investment of engineering resources. Times of day where the intersection utilization exceeds the threshold may still be improved by investment of resources, but may require making some other phase have worse performance to improve the problem phase. Where there is sufficient spare capacity at the intersection, the likelihood is greater that such tradeoffs can be avoided.

Figure 4.15 repeats the split failure plot shown earlier in Figure 4.13, but now the data has been sorted into two categories. The “correctable” split failures indicated by the black bars show split failures occurring in a cycle where the degree of intersection utilization was below a threshold of 0.75. The “not correctable” failures (red bars) occurred during a cycle where the intersection utilization was above this threshold. This gives additional context to the split failure count.

- Most of the failures on phase 2 and 6 are “not correctable” because they occur during peak periods. The overall demand for service across the intersection is high, and making split changes would require prioritization of one phase over another.
- Phases 3, 4, and 5 do have some failures that occur during similar conditions. However, there are more split failures that could be corrected. These tend to occur during off-peak time periods. This indicates an opportunity to fine-tune the splits for the associated timing plans.

Thus, through use of the high resolution data and combining information about the phase status and vehicle detections, there are a number of methods for describing and evaluating the operation. The overall level of utilization for each movement can be quantified in addition to the overall intersection. From this, deficiencies can be discovered, and the most actionable items can be further identified where a deficiency coincides with the presence of enough spare capacity to potentially remedy it.
Figure 4.14. Degree of intersection saturation.

Figure 4.15. Number of split failures per half hour, by phase, sorted by the degree of intersection saturation (using a threshold value of 0.75) into “correctable” and “not correctable” categories.
4.2.5 Utilization of Intersection Capacity under Projected Future Volumes

One additional application of the data is to predict future operating conditions. By applying a growth factor to the observed volumes, it is possible to predict the degree of intersection saturation for future conditions. This helps identify the level of reserve capacity under different possible annual growth rates.

Figure 4.16 shows the degree of intersection saturation at the end of 10 years, based on different potential volumes based on annual growth rates ranging from 0% to 2% increase per year. The 0% rate represents existing conditions. For simplicity, a 10-point moving average is displayed rather than a point cloud. Another way of looking at the data is to examine the amount of spare capacity based on the average value, using the following transformation:

\[ c_s = \max \{0, (1 - X_C)\} \]

Equation 7

Here, \( c_s \) is the spare capacity and \( X_C \) is the degree of intersection saturation. Whenever \( X_C > 1 \), there is no spare capacity left. Figure 4.17 shows a plot of the data viewed from this perspective.

Under the existing conditions (0% growth), the intersection operates during most of the day with 25% or more spare capacity—with the exception of a portion of the AM peak and most of the PM peak. As traffic levels grow, the amount of spare capacity begins to decrease. Under 2% growth over 10 years, the highest rate considered here, the intersection has less than 25% spare capacity for almost the entire day, from 7:00 to 19:00. The intersection has no spare capacity at all during the PM peak.

Another similar analysis examines the number of years of growth at which a particular rate will occur before capacity constraints become a severe issue. In this case, rather than varying the growth rate, the number of years of growth are varied. Figure 4.18 shows the degree of intersection saturation after 2, 4, 6, 8 and 10 years under 2% growth, while Figure 4.19 shows the amount of spare capacity after those time periods. This analysis indicates that, if traffic volumes grow by 2% per year, the midday time periods will have less than 25% spare capacity after about 6 years, while the PM peak will be “maxed out” after 8 years.

The decision to make capital investments across an arterial network would rely on comparing this analysis with other intersections, and reconciling that information with the agency objectives. The results have to be understood in the context of the system. For example, an agency operating an arterial in a dense urban area would likely have to accept high intersection saturation for much of the day for that arterial, while an arterial in a more rural area probably would not. The starting point of the analysis is the understanding of intersection operations at the level of movements.
Figure 4.16. Change in the degree of intersection saturation under different volume growth rates over 10 years.

Figure 4.17. Spare capacity by time of day under different volume growth rates over 10 years.
Figure 4.18. Change in the degree of intersection saturation over time under 2% annual growth.

Figure 4.19. Spare capacity by time of day over time at 2% annual growth.
4.3 **Corridor Level**

As might be apparent from the discussion of introduction-level capacity performance measures, the allocation of capacity across a system is a fairly complex problem with numerous dimensions against which to compare. The dimensions involved in a corridor level analysis are represented conceptually in Figure 4.20. At a single intersection, there are several competing movements (each generally corresponding to a phase), and the performance varies by day of week and time of day. Across a corridor, this range of potential segments of the data is multiplied by the number of intersections.

4.3.1 A Look at Corridor Travel Times

A starting point for corridor level analysis in Module 3 was the travel time measurement along a corridor, and this is also relevant to identifying capacity bottlenecks. Figure 4.21 shows a map of SR 37 in Noblesville, Indiana, which will serve as an example corridor for carrying out an analysis. Figure 4.22 shows a series of CFDs of the travel time across the corridor for weekdays, with the northbound travel time in Figure 4.22a and the southbound travel time in Figure 4.22b. The data is segmented by different times of day using the five timing plans that span the busiest portion of the day. The median (50th percentile) travel times vary between 4 and 6 minutes by time of day in the southbound direction, and between 3.5 and 5 minutes in the northbound direction.

The data shows that the timing plan seems to favor northbound movement for most times of day, yet there are still some timing plans that perform better or worse than other times of day, regardless of direction. The early morning is at the lower end of the range of travel times, which is not surprising given that this is the time of day with the lightest traffic volumes. The PM peak, which has the heaviest volumes, is close to the upper end. However, travel times are actually worse during a few mid-day time periods, specifically the 11–13 and 13–15 time periods, which is unexpected.

![Figure 4.20. Dimensions of analysis for a corridor [4].](image)
Figure 4.21. Map of a four-intersection span of SR 37 in Noblesville, Indiana.
Figure 4.22. Comparison of weekday arterial travel times from different time of day periods.
While travel time data is useful for evaluating the overall system performance, it does not necessarily reveal whether deficiencies in travel time are due to poor progression or due to capacity constraints. The data also only reveals the aggregate performance of the movements included in the travel time route, such as the through movements in this particular case. While it was discovered that some of the midday timing plans have worse performance than might typically be expected, further investigation is needed to determine potential causes and to evaluate all of the movements at the intersections.

4.3.2 Corridor Analysis of Intersection Utilization

A starting point for this analysis is the degree of intersection saturation, which was introduced in the discussion of intersection-level analysis. Figure 4.23 shows a plot of intersection saturation over 24 hours during a representative weekday for the four intersections along SR 37. To simplify the chart, a 10-cycle moving average is presented. The overall trends are similar for all four intersections, with characteristic morning and evening peaks. The relative amount of utilization per intersection is also clear, with Town and Country Blvd. having less utilization than the other three across the entire day. The relative amount of capacity utilization per time period is also apparent. Plan 5, during the PM peak, is the busiest, but plans 3 and 4 have a relatively high amount of utilization at most intersections—higher than the AM peak. This correlates with the higher travel times also occurring during the same time periods.

To better visualize the level of demand at each intersection by time of day, Figure 4.24 presents intersection saturation in each time period as a rank-ordered (Pareto sorted) plot of the individual cycle-by-cycle values. Each sub-plot looks at a particular time period and shows the 60 most congested cycles. For example, the highest intersection saturation occurs during Plan 5 at Greenfield Ave., with a utilization of about 1.25. That time period sees the highest overall level of utilization across the entire corridor. However, as observed earlier, Plans 3 and 4 also have rather high levels of utilization. Sorting the plot in this way enables visualization of the relative level of demand for the different intersections. In particular, Pleasant St. stands out during Plans 3 and 4, while Greenfield Ave. stands out during Plan 5. During Plans 0, 1, and 2, SR 32 has the highest level of utilization.
Figure 4.23. Degree of intersection saturation by intersection (10-cycle moving average).

Figure 4.24. Rank-ordered degree of intersection saturation by intersection.  
(Data originally presented in [1]).
So far, the analysis has considered the time of day and intersection dimensions of the analysis (Figure 4.20). Now it is necessary to consider the individual phases and movements to identify whether operational deficiencies exist in the system. Figure 4.25 shows a series of stacked bar graphs that represent the number of split failures occurring on the individual phases at each intersection, broken down by time of day. Figure 4.25a presents the total number of split failures, while Figure 4.25b shows the number of “correctable” split failures, which are those that occurred during a cycle when there was enough spare capacity to have allowed the demand to have been served, if the green time had been allocated differently. This is determined using a threshold of 0.75 for intersection saturation. The number of split failures that are not correctable are also shown.

The analysis identifies the following five TOD plans and intersections as candidates for improvement:

- Plan 5 sees many split failures across the system. While many of these occur when there is little spare capacity, there are still many remaining split failures that could potentially have been corrected. At Pleasant St. in particular, nearly all of the non-coordinated phases (other than 2 and 6) could benefit from some additional green time during this period.
- Plan 3 stands out as having a high number of correctable split failures, especially at Pleasant St. but also with a modest number at SR 32.
- At Greenfield Ave., most time periods other than plan 5 have a relatively low occurrence of split failures. However, phase 4 stands out during each time of day as a candidate. The other phases are less prominent.
- Similarly, at SR 32, phase 5 has a number of split failures during most times of day that could have been corrected.

Interestingly, phases 2 and 6 rarely experience split failures, according to this analysis. This is likely because, as coordinated phases, these tend to receive most of the green time. Not only are the splits of phases 2 and 6 likely higher because of an objective to emphasize movement along the corridor, but coordination itself tends to favor phases 2 and 6 because the controller has to rest in green on these phases to achieve the scheduled pattern of green bands that facilitate progression. While this is necessary to maintain traffic flow through coordination, it also shows that there are opportunities to fine-tune the splits or other control options to better respond to traffic.
Figure 4.25. Split failures per hour by phase across the corridor. (Data originally presented in [1]).
4.4 Network Level

The network perspective is scoped at assessing the performance of the entire inventory of signalized intersections in an agency’s arterial program. As noted in Module 3, an arterial network containing 100 intersections, with 12 movements per intersection, and seven time-of-day plans across a week, will have about 8,400 different items with a particular level of capacity utilization. Two approaches to identifying overall capacity utilization are presented. In the first, the number of split failures is aggregated across the system to find the intersections with the highest failure rate. In the second, the degree of intersection utilization is compiled across the network to view current and projected rates of intersection capacity utilization, to identify potential opportunities for capital improvements.

4.4.1 Ranking Split Failure Severity Across the Network

Earlier in this module, different performance measures were presented for identifying occurrences of split failures at an intersection. A split failure is a symptom of oversaturation, because there is more demand than can be served by the split during a particular cycle. A high number of split failures is a key identifier of an opportunity to improve signal operations.

The intersection level methodology was applied across the inventory of signals in Indiana that were online and reporting complete data on May 31, 2014 to evaluate the occurrence of split failure by movement and intersection. The type of performance measure used for this purpose was determined by the type of detection.

- Where stop bar detection was in place, the ROR/GOR approach was used. Values of ROR and GOR exceeding 0.80 were used to consider a phase as having a split failure.
- Where advance detection was used, the volume-to-capacity (v/c) ratio approach was used. Cycles with v/c values exceeding 1.00 were considered to be split failures.

Table 4.1 shows a list of the top 30 intersections by rate of split failure occurring during the PM peak (15:00–19:00) on May 31, 2014. Because all of the intersections under consideration operate in an eight-phase scheme, it is convenient to tabulate the data on a phase basis. The data is presented both in terms of the total number of split failures by phase, and in the percentage of cycles in which the phase failed. The list is sorted from greatest to least according to the maximum failure rate of any phase. This brings the intersection with the most severe failure pattern of any phase to the top of the list. For example, had an average value been used rather than the maximum, the top-ranked intersection would have had an average failure rate of only 18% and would have fallen much lower in the ranking.

This analysis shows that phase 6 at SR 37 and SR 144 experienced a failure during every single cycle of the 15:00–19:00 time period, according to the data. The next intersection, US 36 and Steeplechase, had severe failures on two phases. The next step would be to investigate the intersections in detail to determine possible actions that could be taken to improve the quality of service, and then validate whether a positive impact is obtained through a change. The next section of this module contains some case studies where this was done using the high-resolution data.
Table 4.1. Example of agency-wide split failure data: Top 30 intersections, sorted by maximum percentage of failures during the PM peak (1500-1900), May 31, 2014 [5].

<table>
<thead>
<tr>
<th>Signal ID</th>
<th>Location</th>
<th>Number of Cycles</th>
<th>Number of Cycles with Split Failure</th>
<th>Percentage of Cycles with Split Failure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 2 3 4 5 6 7 8</td>
<td>Max 1 2 3 4 5 6 7 8</td>
</tr>
<tr>
<td>4450</td>
<td>SR 37 &amp; SR 144</td>
<td>121</td>
<td>14 0 0 15 7 121 17 1</td>
<td>1 100 12 0 0 12 6 100 14 1</td>
</tr>
<tr>
<td>4417</td>
<td>US 36 &amp; Steeplechase</td>
<td>120</td>
<td>25 116 107 0 2 0 1 10</td>
<td>2 97 21 97 89 0 2 0 1 8</td>
</tr>
<tr>
<td>4464</td>
<td>SR 37 &amp; 141st St</td>
<td>127</td>
<td>0 0 0 8 111 0 4 0</td>
<td>3 87 0 0 0 6 87 0 3 0</td>
</tr>
<tr>
<td>1293</td>
<td>US 31 &amp; Fry Rd</td>
<td>131</td>
<td>6 0 0 25 8 0 69 104</td>
<td>4 79 5 0 0 19 6 53 79</td>
</tr>
<tr>
<td>4428</td>
<td>SR 37 &amp; 146th St</td>
<td>127</td>
<td>17 0 82 0 39 0 10 0</td>
<td>5 65 13 0 65 0 31 0 8 0</td>
</tr>
<tr>
<td>4441</td>
<td>US 36, SR 67 &amp; 42nd St</td>
<td>144</td>
<td>0 0 91 0 0 0 0 0</td>
<td>6 63 0 0 63 0 0 0 0 0</td>
</tr>
<tr>
<td>4437</td>
<td>SR 37 &amp; Southport Rd</td>
<td>120</td>
<td>31 0 30 65 3 0 5 71</td>
<td>7 59 26 0 25 54 3 0 4 56</td>
</tr>
<tr>
<td>399</td>
<td>US 421 &amp; 106th St</td>
<td>144</td>
<td>8 0 30 0 3 82 4 2</td>
<td>8 57 6 0 0 0 2 67 3 1</td>
</tr>
<tr>
<td>2062</td>
<td>US 36 &amp; Raceway Rd.</td>
<td>120</td>
<td>26 0 0 17 61 0 1 4</td>
<td>9 51 22 0 0 14 51 0 1 3</td>
</tr>
<tr>
<td>4469</td>
<td>SR 37 &amp; Smith Valley Rd</td>
<td>120</td>
<td>0 38 0 60 27 0 0 0</td>
<td>10 58 0 32 0 50 23 0 0 0</td>
</tr>
<tr>
<td>4419</td>
<td>US 36 &amp; Ronald Reagan Pkwy</td>
<td>120</td>
<td>36 0 21 33 38 0 59 7</td>
<td>11 49 30 0 18 28 32 0 49 6</td>
</tr>
<tr>
<td>4412</td>
<td>SR 32/38 &amp; SR 37</td>
<td>127</td>
<td>3 11 21 42 49 22 0 0</td>
<td>12 39 2 9 17 33 39 17 0 0</td>
</tr>
<tr>
<td>4399</td>
<td>US 36 &amp; Gable Dr</td>
<td>120</td>
<td>0 0 30 0 1 0 44 0</td>
<td>13 37 0 0 25 0 1 0 37 0</td>
</tr>
<tr>
<td>4408</td>
<td>SR 37 &amp; 126th St</td>
<td>126</td>
<td>10 0 0 14 16 0 42 44</td>
<td>14 35 8 0 0 11 13 0 33 35</td>
</tr>
<tr>
<td>1289</td>
<td>US 31 &amp; Main St</td>
<td>131</td>
<td>0 0 0 10 6 45 12 31</td>
<td>15 34 0 0 0 8 5 34 9 24</td>
</tr>
<tr>
<td>2046</td>
<td>US 31 &amp; County Line Rd</td>
<td>131</td>
<td>41 0 36 10 18 0 5 12</td>
<td>16 31 31 0 27 8 14 0 4 9</td>
</tr>
<tr>
<td>365</td>
<td>US 421 &amp; 96th St</td>
<td>144</td>
<td>0 45 1 9 0 6 0 0</td>
<td>17 31 3 0 31 1 6 0 4 0</td>
</tr>
<tr>
<td>1300</td>
<td>US 31 &amp; Stop 18</td>
<td>122</td>
<td>20 0 29 20 1 0 24 37</td>
<td>18 30 16 0 24 16 1 0 20 30</td>
</tr>
<tr>
<td>1011</td>
<td>US 421 &amp; Retail</td>
<td>144</td>
<td>1 0 41 22 1 0 0 0 19</td>
<td>19 28 1 0 28 15 1 0 0 0</td>
</tr>
<tr>
<td>4419</td>
<td>US 36 &amp; Meijer</td>
<td>120</td>
<td>4 0 0 0 0 8 0 0 34</td>
<td>20 28 3 0 0 7 0 0 28 0 0 0 28</td>
</tr>
<tr>
<td>3677</td>
<td>US 31 &amp; Madison Ave</td>
<td>131</td>
<td>0 0 37 4 0 0 0 0 21</td>
<td>21 28 0 0 28 3 0 0 0 0</td>
</tr>
<tr>
<td>2056</td>
<td>US 421 &amp; I-465 EB Ramp</td>
<td>144</td>
<td>40 0 0 14 0 0 0 0</td>
<td>22 28 28 0 0 10 0 0 0 0</td>
</tr>
<tr>
<td>4415</td>
<td>SR 37 &amp; Pleasant St</td>
<td>126</td>
<td>20 0 0 14 30 0 15 23</td>
<td>23 24 16 0 0 11 24 0 12 18</td>
</tr>
<tr>
<td>4445</td>
<td>US 36, SR 67 &amp; Walmart</td>
<td>144</td>
<td>1 0 0 2 0 0 0 32</td>
<td>24 22 1 0 0 1 0 0 0 0 22</td>
</tr>
<tr>
<td>4493</td>
<td>SR 37 &amp; Pilot</td>
<td>152</td>
<td>0 0 30 0 1 0 0 0 25</td>
<td>25 20 0 0 20 0 1 0 0 0 0</td>
</tr>
<tr>
<td>4430</td>
<td>SR 37 &amp; Greenfield Ave</td>
<td>127</td>
<td>0 0 0 1 25 0 0 0 5</td>
<td>26 20 0 0 0 1 20 0 0 4</td>
</tr>
<tr>
<td>4402</td>
<td>US 36 &amp; Country Club Rd</td>
<td>120</td>
<td>0 0 0 0 0 23 0 0 27</td>
<td>27 19 0 0 0 19 0 0 0 0</td>
</tr>
<tr>
<td>1288</td>
<td>US 31 &amp; Smith Valley Rd</td>
<td>131</td>
<td>2 0 0 1 7 0 12 24</td>
<td>28 18 2 0 0 1 5 0 9 18</td>
</tr>
<tr>
<td>2028</td>
<td>SR 37 &amp; Thompson Rd</td>
<td>172</td>
<td>4 0 29 3 9 0 0 29</td>
<td>29 17 2 0 17 2 5 0 0 0</td>
</tr>
<tr>
<td>4446</td>
<td>US 36, SR 67 &amp; 56th St</td>
<td>144</td>
<td>0 0 24 2 0 0 5 7</td>
<td>30 17 0 0 17 1 0 0 3 5</td>
</tr>
</tbody>
</table>
In addition to identifying oversaturation hot spots throughout the network, it is also important to give an overview of the total system performance. Figure 4.26 shows a plot of the rank-ordered (Pareto sorted) list of failure rates for different times of day throughout the agency inventory. The data in each case was prepared the same way as for Table 4.1, but are sorted from greatest to least value independent of location, which gives an overall profile of the system performance and how it varies by time of day. For example, it is shown that during the midday, PM, and evening time periods, there is at least one phase at one intersection in the system with a 100% failure rate. Additionally, of the five different times of day, the midday actually has a higher rate of split failure than the PM, which might be rather unexpected. The AM peak has fewer split failures than the other time periods, with the exception of the overnight period, in which relatively few failures occur. There are still two intersections, however, that show a relatively high level of phase failure even during the overnight period.

This sort of system profile provides a snapshot of the overall system performance at a moment in time. If this data were tracked over time, it would provide the analyst an ability to evaluate changes in the system and to benchmark the agency’s arterial performance as a whole, from the perspective of efficient capacity allocation, in a similar manner to which travel time data and arrival characteristics would be used to track performance from a smooth traffic flow perspective (as discussed in Module 3).
4.4.2 Ranking Intersections by Capacity Utilization

In addition to the split failures on individual phases, the intersection-level discussion also presented the degree of intersection capacity utilization. This metric was used to evaluate whether any spare capacity existed at the intersection and to classify split failures as correctable or not (that is, whether they might be remedied by adjusting signal timing alone). In addition, the intersection utilization concept was used to understand future intersection performance under projected volumes at different growth rates. At the corridor level, the degree of intersection utilization was helpful for understanding and visualizing the level of demand at different intersections at different times of day.

Similarly, the degree of intersection utilization can be aggregated up to the network level to understand the level of congestion throughout the agency inventory, to project current and future needs of the arterial program. To facilitate this discussion, an analysis was carried out for 115 intersections that were online and reporting high resolution data in the Indiana network on July 1, 2014. A representative value of the intersection capacity utilization was developed based on the observed traffic volumes during the PM peak hour (1700-1800) on that day to estimate the impact on the system during the heaviest time period of traffic loading. After compiling the data for all of the intersections, the resulting list of intersection values were sorted into five bins and plotted to show their geographic distribution.

Figure 4.27 shows the resulting network map, showing the geographic distribution of the intersections across the state and the degree of intersection saturation. A more detailed view of the Indianapolis area is provided in Figure 4.28. From this representation of the performance measure, it is possible to quickly discern which intersections currently face high levels of capacity utilization and which have lower levels. In Figure 4.28, for example, there are two intersections in the highest group on US 36 West and one in SR 37 North. US 52, on the other hand, lies entirely within the lowest groups. US 36 mostly lies in the middle group but there are a few intersections with high utilization along that corridor.
Figure 4.27. Agency-wide analysis: Degree of intersection saturation ($X_c$) for the PM peak hour (1700-1800). Data is shown for 115 intersections with data from July 1, 2014.
Figure 4.28. Agency-wide analysis: Degree of intersection saturation ($X_c$) for the PM peak hour (1700-1800). Close-up of the Indianapolis area.
Similar to how the rate of split failure was used to rank the intersections to develop a system profile, the degree of intersection capacity utilization can also be used as a basis for such a ranking. In addition, by adjusting the component volumes, it is possible to project the possible future system state based on assumed growth rates.

Figure 4.29 shows a rank-ordered (Pareto sorted) list of the 115 intersections included in the survey from the highest to the lowest degree of intersection utilization during the peak period on the sample date, along with that value re-calculated based on volumes adjusted for different growth rates over a 10-year period. The 0% growth rate represents the current conditions.

At present, there are about 19 intersections out of the 115 currently operating with a utilization value of more than 1.0 during the PM peak. This represents the number of intersections currently lacking spare capacity to handle additional demand. The highest value of all is 1.43. After 10 years of 2% annual growth in the volumes, the number of intersections above 1.0 would increase from 19 to 35, and the highest value would increase from 1.43 to 1.75. At more moderate annual growth levels, the increase is less. The analysis shows that a rather small change in the annual growth rate can lead to a substantial increase in the number of intersections meeting or exceeding capacity.

An alternative use of growth rates would be to project the number of years before the system reaches a certain performance level. Figure 4.30 shows a rank-ordered list with projects according to different numbers of years at 2% growth, ranging from current conditions (0 years) to 10 years. For example, the current number of intersections at capacity is 19. This can be expected to surpass 30 intersections within 6 years if the growth rate is 2%.
Figure 4.29. Agency-wide analysis: Pareto sorted list of intersections by degree of intersection saturation (based on afternoon peak hour volumes), under different projected annual growth rates after 10 years.

Figure 4.30. Agency-wide analysis: Pareto sorted list of intersections by degree of intersection saturation (based on afternoon peak hour volumes), under different number of years at 2% annual growth.
4.5 Guidance for Operational Analysis

This module concludes by presenting a few case studies where the high resolution signal event data was applied to analyze arterial operations. This includes three case studies in which timing plans were adjusted and the performance of the system was tracked in response to the changes.

- The first case study considers a split change that re-assigns green time from one group of phases to another group of phases.
- The second case study considers a change to the signal operation affecting all phases, by enabling a feature that allows green time to be more dynamically re-allocated.
- The third case study examines the impact of cycle length changes on a corridor in terms of both capacity utilization and quality of progression.

4.5.1 Measuring the Impact of a Split Adjustment

One of the simplest signal timing changes that can be made is an adjustment of the splits, which controls the distribution of green times to the various phases at an intersection. This would be a fairly typical initial approach to eliminating split failures on a phase or to address complaint calls.

In this case study, the intersection of US 31 and 126th St. in Carmel, Indiana was checked to determine the presence of split failures. Figure 4.31 shows an overview of this intersection including its control sequence, which follows a typical eight-phase pattern and detection layout. Note that there are stop bar detection zones on all of the lanes except for the northbound and southbound through and right turn lanes, which have only advance detectors.

The detector layout determined the type of performance measures that would be used to assess the performance of the splits. All of the phases whose lanes featured stop bar detection would use occupancy based metrics (ROR versus GOR), while those phases that were served by advance detection instead used a count based metric, the volume-to-capacity (v/c) ratio.

The results of the initial field analysis are shown in Figure 4.32. The occupancy based metrics are shown in Figure 4.32a, with ROR plotted against GOR, arranged in an eight-phase format. Phases 2 and 6 had no stop bar detection, so this performance measure is not available for them. For the others, the number of split failures is indicated by the callout; this corresponds to the number of cycles for which both ROR and GOR exceeded a threshold value of 0.8. Figure 4.32b and Figure 4.32c respectively illustrate the v/c ratio over the same time period for phases 2 and 6.

Scanning across the ROR/GOR plot, it is clear that phases 3 and 4 have the highest number of split failures, with phases 1 and 8 also having a fair amount. Phases 5 and 7 exhibit very few split failures. Also, phases 2 and 6 have rather low v/c ratios throughout the entire time period, with the average values just above 50% and the cycle-by-cycle values generally less than 75%.
Figure 4.31. US 31 and 126th St. [3].
(a) Eight-phase ROR versus GOR plots with number of split failures indicated. Data is tabulated for the 9:00–15:00 time period.

(b) Phase 2 v/c ratio

(c) Phase 6 v/c ratio

Figure 4.32. Capacity utilization performance measures for an example intersection [3].
Based on this analysis, the agency decided to implement a split adjustment to relieve some of the split failures on the phase with the highest number, phase 3. The split adjustments are illustrated graphically in Figure 4.33. A split time of 4% of the cycle was transferred from phase 2 to phase 3, increasing its split from 16% to 20%. Because this time moved from the mainline phase group to the cross street phase group (that is, across the barrier in Figure 4.33), it was necessary to accompany this adjustment to the first ring’s splits with another in the second ring. Because phase 8 had the most split failures in the second ring, 4% was transferred from phase 6 to phase 8.

![Figure 4.33. Split adjustments at US 31 and 126th St [3].](image)

The outcomes of the change are shown in Figure 4.34. This graph shows the count of the split failures, as determined by cycles where the ROR and GOR were both above 0.8, on each phase before and after the adjustment. As might be expected, since phases 3 and 8 saw an increase in their split times, both phases saw a decrease in the number of split failures; the reduction in both cases was more than half. Phase 4, which did not see any increase in the programmed splits, did not see a reduction in the number of split failures. Similarly, phases 5 and 7 saw no changes, and seemed to have few problems both before and after the split adjustment. Interestingly, phase 1 saw a slight increase in the number of split failures; this was most likely due to variation in the volume, since the signal timing did not change on that phase.

The impacts on phases 2 and 6, which both gave up green time, are explained by Figure 4.35, which shows the v/c ratios for the two phases before and after the adjustment. Figure 4.35a and Figure 4.35b respectively show the before and after v/c ratios for phase 2, while Figure 4.35c and Figure 4.35d show the before and after v/c ratios for phase 6. There was a very slight increase in the average v/c ratios for both phase 2 and 6, which is what would be anticipated with a decrease in the green time. However, the average v/c ratios still fall below 60% for both phases.
Figure 4.34. Number of split failures before and after split adjustment [3].

Figure 4.35. Volume-to-capacity ratios before and after a split adjustment [3].
Because phases 2 and 6 also affect coordination, it was important to check whether the split adjustment had any effect on this aspect of the performance. PCDs showing the operation before and after the split adjustment are shown in Figure 4.36. Phase 2’s before and after performance is shown respectively by Figure 4.36a and Figure 4.36b, while phase 6 is examined in Figure 4.36c and Figure 4.36d. In both pairs of before/after graphics, very little difference is visually apparent between the “before” and the “after” data. If examined closely, evidence of the decrease in the green time can be seen in a very slight narrowing of the green band. However, there is minimal impact to progression; the relationship between the arrivals and the green band are virtually identical for both northbound and southbound traffic flow.

This case study demonstrates the use of performance measures to validate one of the simplest signal timing adjustments. The performance measures were used to identify a deficiency, and a potential remedy was suggested from that analysis by identifying which phases needed capacity and which phases could give up capacity. After implementing the change, it was possible to validate the impact and verify that the adjusted splits had the desired effect. Furthermore, it was possible to verify that the adjustment did not have an undesirable effect on other objectives, namely that the quality of progression was not disrupted.
Figure 4.36. PCDs for time periods before and after the offset adjustment, for both directions.
4.5.2 Measuring the Impact of an Intersection Control Policy Change

The next case study examines a more critical change to the green time allocation strategy at an intersection. An advanced controller feature that enables capacity to be reallocated in a more dynamic fashion was implemented and tested using the high-resolution data. The feature in question is referred to by different names depending on the vendor, but some of the descriptions are “early yield” or “actuated-coordinated” operation. It refers to actuation of the coordinated phase.

In coordination, the controller is required to serve the coordinated phases at a scheduled time. The way this is implemented is that, in each cycle, each non-coordinated phase receives an opportunity for service. During a particular window of time\(^3\), if there is a call, the controller will serve the phase. After that window closes, the phase cannot be served until the next cycle. Each non-coordinated phase also must terminate by a required time. This forces the controller to get back into green for the coordinated phases on schedule. Afterward, the controller stays in green on those phases, and is not allowed to be released back to the non-coordinated phases until after it reaches a time in the cycle called the “yield point.”

With actuation, the non-coordinated phases often end early, which allows them to forward the balance of their time to phases that come later. Because the coordinated phases are not allowed to end early, they typically end up absorbing all of this “extra” green time whether they need it or not. The impact is sometimes noticeable, for example by drivers or pedestrians waiting on the cross street for the coordinated green to end (sometimes in absence of any traffic).

The early yield strategy moves the yield point to an earlier time in the cycle, creating an actuated portion of the coordinated phases. The difference in operation is explained schematically by Figure 4.37. Note that in Figure 4.37a, phases 2 and 6 are entirely “non-actuated,” and the yield point occurs at the end of the phases. In Figure 4.37b, the yield point has been moved earlier in the cycle, establishing an “actuated” portion of phases 2 and 6. Under this scheme, if vehicle detections extend phases 2 and 6, they will continue to be served. However, if the detectors observe a gap in the traffic, the controller will terminate those phases and move on to the others. This allows the coordinated phases to give up green time to the non-coordinated phases, increasing the pool of spare capacity that can be utilized by subsequent phases. The impact of this strategy has been shown to be equivalent to a split change [6], with the difference that the reallocation occurs dynamically rather than according to the specific splits selected to relinquish and inherit cycle time.

---
\(^3\) Called a “permissive period” or “permissive window.”
(a) Conventional coordinated operation.

(b) Coordinated operation with early yield.

Figure 4.37. Explanation of the operational change.
High-resolution data was obtained during a period before and after implementation. In this study [7], because high quality count data was available at the intersection, the v/c ratio was selected as the basis for the analysis. Figure 4.38 presents CFDs of the v/c ratio per phase for representative dates in the before and after time periods, for the AM peak (Figure 4.38a) and the PM peak (Figure 4.38b). Note that the placement of phases 5 and 6 are swapped to reflect the lagging left turn condition in the phase sequence. In addition to the graphical shift in the CFDs, the distributions of v/c ratios were used to determine whether the changes were statistically significant at a 90% confidence level.

In both the AM and PM, phases 3 and 5 benefit strongly from the strategy change. These phases immediately follow the coordinated phases (2 and 6), and directly benefit from the yielding of time. In the AM peak (Figure 4.38a), phases 7 and 8 also benefit slightly (although those changes are not statistically significant). There is an increase in the v/c ratio for coordinated phase 6 in the PM peak (Figure 4.38b), which is not unexpected since the green time is being reduced. Coordinated phase 2 has almost no difference in the v/c ratio in both the AM and the PM.

As in the split adjustment case study, the progression objective is also affected by changes to green times on the mainline. To verify whether the new strategy impacted progression, PCDs from before and after the change were constructed. Figure 4.39a shows a PCD for the northbound movement before the change, while Figure 4.39b shows conditions afterward. The before data shows evidence of wasted green time in the fact that few vehicles are served in the later portion of the green (Figure 4.39a, callouts “i” and “ii”). This observation corresponds to the driver waiting on the side street when there is no traffic on the mainline. The post-implementation data shows that this portion of time has been eliminated (Figure 4.39b, callouts “iii” and “iv”). Note also that the end of green time now varies around an expected value, rather than always being exactly the same time in cycle. During the AM and PM peaks, progression is still quite successful, with most vehicles arriving on green. There is, however, some evidence of an opportunity to refine the offset during some of the midday time periods, as some of the platoons are cut off during green (Figure 4.39b, callout “v”).

This case study demonstrated a signal timing change where the operational strategy was itself significantly altered, affecting all phases. It was not known prior to the change what the impact might be, but analysis of the data was able to reveal in detail what took place. The dynamic rebalancing of splits with the implementation of early yield was seen to shift the degree of utilization of many phases, reducing split failures for a handful of phases. A view of the PCDs revealed the cyclic nature of the arrivals and how the green time was used more efficiently after the feature was enabled by eliminating periods of wasted green time.
Figure 4.38. Graphs of cumulative frequency of V/C ratios. Arrows indicate change in the average value [7].
(a) Typical coordination with non-actuated coordinated phase.

(b) Coordination with actuated-coordinated phase.

Figure 4.39. Impact of early yield on progression [8].
4.5.3 Investigating the Impact of Cycle Length Changes

This third case study investigates the impact of another change to signal timing that profoundly influences the traffic control on a corridor: the cycle length of the coordination plan.

In this study, the agency desired to know whether a signal timing policy, which emphasized finding the “right” cycle length for a corridor and fine-tuning offsets and splits around that parameter, was an effective strategy to establish signal timing. Experience had shown that, during periods of oversaturation, there were diminishing returns in increasing the cycle length to improve throughput [9]. It was also desired to explore the impact of cycle length on other aspects of operation such as the occurrence of split failures and the quality of progression. Therefore, a study [10] was carried out in which the cycle length for an entire corridor was varied through a 20-second range to identify the changes that occurred.

Figure 4.40 shows a map of the five-intersection corridor including the distances between intersections and the locations of two vehicle sensors (“Bluetooth Locations”) on the endpoints of the corridor. Data collection took place over several weeks; during this time, each cycle length was tested over two weekdays. On day 1, the cycle length was adjusted from its former value but none of the other settings were changed. On day 2, the offsets were optimized based on data from day 1, so that conditions under optimal offsets could be identified. The cycle lengths were adjusted in a random order to avoid any correlation with the time period.

The outcome for the capacity utilization is presented in Figure 4.41, which shows the proportion of split failures occurring in the system. In this study, the “gap out” condition, in which a phase terminates early due to a gap in the detector occupancy, was used to estimate whether or not a phase was likely to have had a failure. Figure 4.41 shows that, as the cycle length was increased from 104 to 124 seconds, the proportion of phases that were likely to have experienced split failures continually decreased from about 40% to less than 30%.

The impact on corridor travel times is illustrated by Figure 4.42. This plot shows the median value and the 25th and 75th percentiles of the measured travel times under each different tested cycle length. The results show that, as cycle length increases, the travel time slightly increases. The trend is stronger for the southbound direction; it is more ambiguous for northbound traffic. The increase in travel time can be attributed to the fact that, when vehicles do have to stop, they must wait for a longer amount of time (when cycle length increases) before the next start of green.
Figure 4.40. A five-section span of SR 37 [10].
Figure 4.41. Percentage of phases forced off at different cycle lengths [10].

Figure 4.42. Travel time distributions[10].
The conclusion from this study was that there was no particular cycle length within the range that yielded a superior performance compared to the others. Increasing the cycle length was shown to lead to a decrease in the number of likely split failures, with an increasing number of phases terminating early. At the same time, the corridor travel times increased, which shows that there is likely a performance tradeoff between the two objectives of local and corridor operation.

This final case study showed results from a large scale corridor study in which a number of changes were made to the system over a period of time. During the analysis, the data was able to provide rapid feedback which enabled optimal offsets to be determined for each cycle length adjustment within one day after implementation. The scale of change is similar to the type of fine tuning that might follow the deployment of a new timing plan or a new control scheme. However, in this case, the analysis was done with very little time spent manually observing traffic. The performance measure data provides feedback on the process that would normally only be obtainable through field observations.
4.6 Summary

This module presented a series of performance measures related to the measurement of intersection capacity and demand for the analysis of arterial networks. These were applied to analysis use cases at the intersection, corridor, and network levels.

- At the intersection level, performance measures for quantifying the level of capacity utilization on a given phase within a given cycle were introduced. Depending on the type of detection available, either occupancy or count based performance measures could be developed. Both metrics provided a means of evaluating whether there were operational deficiencies for a particular movement at an intersection. Furthermore, compiling the data across all of the competing movements at the intersection enabled the overall level of intersection utilization to be appraised. By comparing the occurrence of split failure with the level of intersection utilization, the number of “correctable” split failures could be identified. Also, by adjusting the traffic volumes going into this calculation, future levels of utilization could be predicted for different growth rates.

- At the corridor level, system performance could be identified by aggregation of the intersection level performance measures. Visualization of the results could be adjusted to match the use case under consideration. In this discussion, two such views were included. One was a rank-ordered sorting of the observed levels of intersection utilization, which enables the overall demand at each intersection to be identified relative to the other intersections in the corridor. The data was segmented by time of day to show variation from one timing plan to another. The other visualization showed the aggregated number of split failures segmented by phase, time of day, and intersection. This facilitated the identification of trends across those three dimensions.

- At the network level, the analysis consisted of further aggregating the intersection-level performance measures into a larger pool of intersections for comparison with one another. The number of split failures and the degree of intersection saturation were used as different points of comparison, which correspond to different objectives. The former allows the analyst to identify the movements with the most severe split failures in the arterial network. The latter provides a profile of the entire system. Repeating the adjustment of background volumes at the network level shows how the system performance would likely evolve under different traffic growth rates.

The performance measures were demonstrated in a series of case studies that applied them to measure the impact of signal timing changes. These ranged in complexity from a simple split adjustment to a range of cycle length adjustments. The objective of establishing equitable and efficient capacity allocation was coupled with an objective to maintain smooth traffic flow and a high quality of progression along the corridor. In each case, the impact on the system could be quantified by a combination of performance measures aimed at each of the competing objectives. In the first two case studies, an improvement to the local control came at little to no cost to the quality of progression. In the final case study, a decrease in likely split failures was observed at the intersections in the corridor, while the travel times through the corridor gradually increased.
4.7 References


The “Small Business Innovation Development Act of 1982” (Pub. L. No. 97-219), along with reauthorizing legislation (Pub. L. No. 99-443 and Pub. L. No. 102-564, the “Small Business Research and Development Enhancement Act of 1992”), seeks to encourage the initiative of the private sector and to use small business effectively to meet federal research and development objectives. To comply with statutory obligations of the Act, the U.S. Department of Transportation established the Small Business Innovation Research (SBIR) Program, which conforms to the guidelines and regulations provided by the Small Business Administration. Annually, small businesses are solicited to submit innovative research proposals that address the high-priority requirements of the U.S. Department of Transportation and that have potential for commercialization.

This report was developed through a partnership between Traffax, Inc., and Purdue University with funding from a Phase III SBIR contract (DTFH6114C00035) with the Federal Highway Administration. The project, entitled “Sensor Fusion and MOE Development for Off-Line Traffic Analysis of Real Time Data,” created and refined methods and tools for the characterization of performance along arterial corridors.

This report is part of a series of reports published in collaboration with USDOT, Traffax, Inc., and Purdue University. The full report series is available for download at http://docs.lib.purdue.edu/apmtp/.

The Indiana legislature established the Joint Highway Research Project in 1937. In 1997, this collaborative venture between the Indiana Department of Transportation and Purdue University was renamed as the Joint Transportation Research Program (JTRP) to reflect state and national efforts to integrate the management and operation of various transportation modes. Since 1937, the JTRP program has published over 1,600 technical reports. In 2010, the JTRP partnered with the Purdue University Libraries to incorporate these technical reports in the University’s open access digital repository and to develop production processes for rapidly disseminating new research reports via this repository. Affiliated publications have also recently been added to the collection. As of 2017, the JTRP collection had over 1.5 million downloads, with some particularly popular reports having over 20,000 downloads.