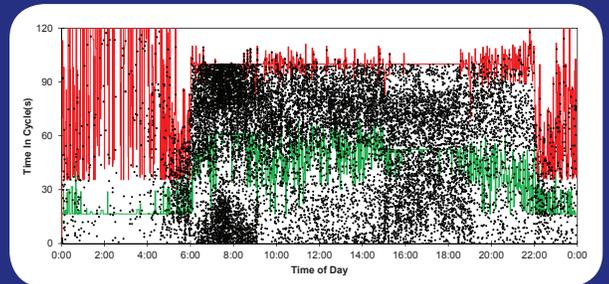


PERFORMANCE MEASURES FOR TRAFFIC SIGNAL SYSTEMS

An Outcome-Oriented Approach



Christopher M. Day, Darcy M. Bullock, Howell Li, Stephen M. Remias, Alexander M. Hainen, Richard S. Freije, Amanda L. Stevens, James R. Sturdevant, and Thomas M. Brennan



PERFORMANCE MEASURES FOR TRAFFIC SIGNAL SYSTEMS

An Outcome-Oriented Approach

*Christopher M. Day, Darcy M. Bullock, Howell Li, Stephen M. Remias, Alexander M. Hainen,
Richard S. Freije, Amanda L. Stevens, James R. Sturdevant, and Thomas M. Brennan*

Recommended Citation

Day, C. M., D. M. Bullock, H. Li, S. M. Remias, A. M. Hainen, R. S. Freije, A. L. Stevens, J. R. Sturdevant, and T. M. Brennan. *Performance Measures for Traffic Signal Systems: An Outcome-Oriented Approach*. Purdue University, West Lafayette, Indiana, 2014. doi: 10.5703/1288284315333.

Acknowledgments

This work was supported in part by the Pooled Fund Study TPF-5(258) led by the Indiana Department of Transportation (INDOT) and supported by the state transportation agencies of California, Georgia, Minnesota, Mississippi, New Hampshire, Pennsylvania, Texas, Utah, and Wisconsin, the Federal Highway Administration Office of Operations and Resource Center, and the Chicago Department of Transportation. This monograph summarizes performance measures developed by the Pooled Fund project and material developed previously by projects funded by the Joint Transportation Research Program, Traffax/USDOT (Project No. DTRT57-11-C-10009), and the National Cooperative Highway Research Program Project 3-79a. The authors also wish to thank the many agencies and vendors we have collaborated with during the development of data logging software and the infrastructure for harvesting the data.

The contents of this monograph 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.

Copyright

Copyright 2014 by Purdue University. All rights reserved.

Print ISBN: 978-1-62260-280-3

ePUB ISBN: 978-1-62260-296-4

ABSTRACT

PERFORMANCE MEASURES FOR TRAFFIC SIGNAL SYSTEMS: AN OUTCOME-ORIENTED APPROACH

This monograph is a synthesis of research carried out on traffic signal performance measures based on high-resolution controller event data, assembled into a methodology for performance evaluation of traffic signal systems. High-resolution data consist of a log of discrete events such as changes in detector and signal phase states. A discussion is provided on the collection and management of the signal event data and on the necessary infrastructure to collect these data. A portfolio of performance measures is then presented, focusing on several different topics under the umbrella of traffic signal systems operation. System maintenance and asset management is one focus. Another focus is signal operations, considered from the perspectives of vehicle capacity allocation and vehicle progression. Performance measures are also presented for nonvehicle modes, including pedestrians, and modes that require signal preemption and priority features. Finally, the use of travel time data is demonstrated for evaluating system operations and assessing the impact of signal retiming activities.

CONTENTS

1. INTRODUCTION	1
1.1 System Perspective: Processes and Stakeholders	1
1.2 Physical Elements of a Traffic Signal System	4
2. INFRASTRUCTURE REQUIREMENTS.	6
2.1 Vehicle Detectors.	6
2.2 Data Acquisition	8
2.3 Communications	9
2.4 Data Infrastructure	11
2.5 Data Size	12
2.6 Performance Measure Implementations	13
3. ASPECTS OF SIGNAL TIMING.	16
3.1 Types of Signal Operation	16
3.2 Control Elements.	16
3.3 Linking Vehicle Detection and Signal Output.	19
3.4 Signal Cycles.	21
3.5 Actuation	22
3.6 Signal Coordination.	24
3.7 Preemption and Priority	27
4. DATA METHODOLOGY	32
4.1 Performance Measure Methodology Overview	32
4.2 Data Specification	33
4.3 Demonstration	38
4.4 Case Study	38
5. CAPACITY PERFORMANCE MEASURES	41
5.1 Traffic Volume Visualization	41
5.2 Capacity Utilization Concepts.	44
5.3 Cycle Length.	47
5.4 Green Time and Capacity Allocation.	49
5.5 Volume and Capacity Utilization	51
5.6 Capacity Allocation and Utilization for Protected-Permitted Phases	55
5.7 Phase Termination.	58
5.8 Green Occupancy Ratio	59
5.9 Green Occupancy Ratio and Red Occupancy Ratio	59
5.10 Degree of Intersection Saturation	62
6. PROGRESSION PERFORMANCE MEASURES	67
6.1 Vehicle Delay Definitions	67
6.2 Delay and Quality of Progression	68
6.3 Progression Performance Measures	71
6.4 Delay Estimates from Measured Arrival Profiles	73
6.5 Purdue Coordination Diagram	75
6.6 Flow Profiles.	79
6.7 Characterizing Platoon Formation and Dispersion	84
6.8 Maximum Queue Length from Shockwave Estimation	84
7. MULTIMODAL PERFORMANCE MEASURES.	91
7.1 Pedestrian Demand	91
7.2 Preemption and Priority	92
8. MAINTENANCE PERFORMANCE MEASURES	96
8.1 Communications Quality	96
8.2 Data Completeness	97
8.3 Detector Failures.	99

9. OUTCOME ASSESSMENT	100
9.1 Travel Time Data Sources	100
9.2 Data Processing and Analysis	101
9.3 Comparison with Crowdsourced Data	101
9.4 Sensor Placement and Data Reduction Techniques	103
9.4.1 Origin-Based Travel Time.	103
9.4.2 Destination-Based Travel Time	103
9.4.3 Subsegment Travel Time, or Control Delay	103
9.5 Commercial Probe Data	106
9.6 Cost Assessment of a Corridor Improvement	108
9.7 Signal Retiming Aging	110
9.7.1 Weekly Analysis of Signal Retiming Aging	110
9.7.2 Monthly Analysis of Signal Retiming Aging	110
10. SUMMARY	114
10.1 Overview of Signal Operations and High-Resolution Data	114
10.2 Performance Measures	114
10.3 Future Opportunities	114
REFERENCES	115
APPENDIX. Required Data Elements for Performance Measures	119

LIST OF TABLES

Table	Page
Table 3.1 Operating modes for signal systems broken down by local control and system control options	16
Table 3.2 Phase timing parameters in various controllers	25
Table 3.3 Coordination reference points for selected controllers	31
Table 4.1 Active phase events	34
Table 4.2 Active pedestrian phase events	34
Table 4.3 Barrier and ring events	34
Table 4.4 Phase control events	35
Table 4.5 Overlap events	35
Table 4.6 Detector events	36
Table 4.7 Preemption events	36
Table 4.8 Coordination events	37
Table 4.9 Cabinet/system events	38
Table 4.10 Excerpt from data table	39
Table 5.1 Capacity parameters from field observation	47
Table 5.2 Recommended base saturation flow rates (vehicles per hour per lane)	47
Table 5.3 Yield point cycle length calculation	48
Table 5.4 Real-time cycle length calculation	49
Table 5.5 Cycle length data table	50
Table 5.6 Calculation of effective red and green (phase 2)	51
Table 5.7 Green time data table (phase 2)	51
Table 5.8 Example vehicle count data (phase 2)	54
Table 5.9 Example cycle-by-cycle vehicle count and equivalent hourly volume (phase 2)	55
Table 5.10 Example data table for volume-to-capacity ratio calculation (phase 2)	55
Table 5.11 Protected and permitted capacity calculation. Volumes and capacities are shown as number of vehicles	60
Table 5.12 X_C calculation, phase group 1	65
Table 5.13 X_C calculation, phase group 2	66
Table 5.14 Final steps in the calculation of X_C	66
Table 6.1 Arrival type definition table, based on HCM Exhibit 15-4	72
Table 6.2 Example data table for POG and AT calculation	73
Table 6.3 Example calculation of input-output delay for two service instances of phase 2 on October 17, 2012	77
Table 6.4 Using high-resolution event data to determine departure and arrival shockwaves	90
Table 9.1 Median travel time savings on US-31 in Kokomo, Indiana	109
Table 10.1 Estimated cost of deploying controller-based high resolution data collection	114

LIST OF FIGURES

Figure	Page
Figure 1.1 Hierarchy of needs in traffic signal system operations and maintenance	1
Figure 1.2 Context for signal system performance measures	2
Figure 1.3 Parameter design/adjustment process	3
Figure 1.4 Maintenance process	3
Figure 1.5 Equipment procurement process	4
Figure 1.6 Overview of elements in a traffic signal system	5
Figure 2.1 Presence versus count (pulse) detection	7
Figure 2.2 Potential detector locations on a typical signal approach	7
Figure 2.3 Pendleton Pike and 56th Street, Indianapolis, Indiana	8
Figure 2.4 Example frame from video for visual counting	9
Figure 2.5 Comparison of visual counts versus loop detector counts	10
Figure 2.6 Communications architecture	11
Figure 2.7 Deployment of automatic data downloader installed on a compact Linux computer	11
Figure 2.8 Data infrastructure (simplified example)	12
Figure 2.9 Horizontal table partitioning	13
Figure 2.10 Data storage requirements over time for one intersection	13
Figure 2.11 Indiana Department of Transportation performance measures website	14
Figure 2.12 Elkhart County Centrac system	14
Figure 2.13 Utah Department of Transportation performance measures website	15
Figure 3.1 Events in local signal phase timing	17
Figure 3.2 Basic observable events for monitoring phase performance	18
Figure 3.3 Standard configuration for dual-ring, eight-phase operation (north-south major road)	18
Figure 3.4 Example overlaps for a right-turn arrow indication	19
Figure 3.5 Example of a negative overlap	19
Figure 3.6 Counting intervals	20
Figure 3.7 Permitted left-turn interval definitions	21
Figure 3.8 Permitted left-turn definitions for five-section head configuration	21
Figure 3.9 Relationship between service instances and cycles	22
Figure 3.10 Variable gap timing	22
Figure 3.11 Added initial timing	23
Figure 3.12 Phase maximum timing (fully actuated operations)	23
Figure 3.13 Dynamic maximum timing	24
Figure 3.14 Signal coordination concepts	27
Figure 3.15 Phase split timing (noncoordinated phase)	28
Figure 3.16 Phase split timing (coordinated phase)	29
Figure 3.17 Actuated-coordinated phases	30
Figure 3.18 Coordination reference points	31
Figure 3.19 Common preemption events	32
Figure 3.20 Events specific to railroad preemption	32

Figure 3.21 Transit signal priority	32
Figure 4.1 Performance measure analysis workflow	33
Figure 4.2 US 36 (Pendleton Pike) in Indianapolis, Indiana	40
Figure 4.3 Pendleton Pike and Post Road	41
Figure 5.1 15-minute vehicle volumes by lane	42
Figure 5.2 Vehicle volume by phase (vehicles per hour)	43
Figure 5.3 Vehicle volume by phase (vehicles per hour per lane)	43
Figure 5.4 Corridor peak hour volumes (a.m.)	44
Figure 5.5 Corridor peak hour volumes (p.m.)	45
Figure 5.6 Theoretical utilization of capacity by a fully saturated movement	46
Figure 5.7 Background cycle length	48
Figure 5.8 Effective cycle length	48
Figure 5.9 Plot of cycle length at two intersections where pattern transitions at one intersection occur one hour earlier (hypothetical example of a time zone error)	49
Figure 5.10 Cycle length under different control types	50
Figure 5.11 Green time (s)	51
Figure 5.12 Green time (s), eight phases	52
Figure 5.13 Phase 2 capacity (veh/h)	52
Figure 5.14 Capacity (veh/h), eight phases	53
Figure 5.15 g/C ratio	53
Figure 5.16 g/C ratio, eight phases	54
Figure 5.17 Vehicle count	55
Figure 5.18 Vehicle count, eight phases	56
Figure 5.19 Equivalent hourly volume	56
Figure 5.20 Equivalent hourly volume	57
Figure 5.21 Volume-to-capacity ratio	57
Figure 5.22 Volume-to-capacity ratio, eight phases	58
Figure 5.23 Through phases related to capacity calculation for phase 3	58
Figure 5.24 Phase 3 capacity (veh/h): Protected and permitted components	59
Figure 5.25 Distributions of reasons for phase termination by half-hour bin	61
Figure 5.26 Definitions of barrier crossing times	61
Figure 5.27 Phase termination diagram showing time redistribution opportunities	62
Figure 5.28 Phase 3 green occupancy ratio	62
Figure 5.29 Phase 3 volume-to-capacity ratio	63
Figure 5.30 The intersection of US 31 and 126th Street in Carmel, Indiana	63
Figure 5.31 ROR_5 versus GOR plot for phase 4 at US 31 and 126th Street	64
Figure 5.32 ROR_5 versus GOR plots for eight phases, and volume-to-capacity ratio for phases 2 and 6	64
Figure 5.33 Explanation of critical path in eight-phase, dual-ring signal operation	65
Figure 5.34 Degree of intersection saturation, X_c	67
Figure 6.1 Delay definitions, with reference to the perspective of an advanced detector	68
Figure 6.2 Delay and traffic flow characteristics with uniform arrivals: (a) time-space diagram; (b) queue profile	69

Figure 6.3 Delay and traffic flow characteristics with platoon arrivals: (a) time-space diagram; (b) queue profile	70
Figure 6.4 Relationship between percent on green, delay, and offset	71
Figure 6.5 Percent on green	73
Figure 6.6 Arrival type	74
Figure 6.7 Obtaining arrival and departure profiles from field data	74
Figure 6.8 Input-output delay polygons	75
Figure 6.9 Concepts for “input-output” delay estimation	76
Figure 6.10 Estimated delay (input-output method)—calculation example	78
Figure 6.11 Estimated delay (input-output method)	78
Figure 6.12 Coordination diagram for one cycle	79
Figure 6.13 Coordination diagram for a 30-minute period	79
Figure 6.14 Coordination diagram for 24 hours: northbound phase 6, SR 37 and Greenfield Avenue, July 25, 2009	80
Figure 6.15 Coordination diagram for 24 hours: westbound phase 2, US 36 and Post Road, October 17, 2012	80
Figure 6.16 Coordination diagram for adaptive control in a simulation environment	80
Figure 6.17 Vehicle origins and destinations on a link	81
Figure 6.18 Using upstream phase status to determine vehicle origins	81
Figure 6.19 Coordination diagram for the southbound through movement at a diamond interchange (I-465 and SR 37 South in Indianapolis, Indiana) for Wednesday, June 5, 2013	81
Figure 6.20 Creation of a cyclic flow profile	82
Figure 6.21 PCD with detail on the 06:00–18:00 time period	83
Figure 6.22 Flow profile (07:00–08:00)	84
Figure 6.23 Flow profiles by hour	85
Figure 6.24 Finding downstream vehicles that are likely to have departed from the upstream intersection through the coordinated movement	86
Figure 6.25 Measured platoons based on upstream beginning of green during fully actuated, noncoordinated operation: SR 37 in Noblesville, Indiana, June 30, 2010, 22:00–24:00	87
Figure 6.26 Matching measured platoons against modeled platoons: example results from SR 37, for data from March 12, 2011, 14:00–16:00	88
Figure 6.27 Queuing shockwave approaching advance detector	89
Figure 6.28 Queue occupying advance detector as phase turns green and more vehicles arrive	89
Figure 6.29 Saturated flow departures reaching the maximum back of queue	89
Figure 6.30 Flow rate change once queue is discharged past advance detector	90
Figure 6.31 Estimated queue lengths	91
Figure 7.1 Northwestern Avenue and Stadium Drive, West Lafayette, Indiana	92
Figure 7.2 Percentage of cycles with pedestrian phases at Northwestern and Stadium, September 19, 2007	93
Figure 7.3 Actuation-to-service time, phase 4, Northwestern and Stadium, September 19, 2007	93
Figure 7.4 Conflicting volume, phase 4, Northwestern and Stadium, September 19, 2007	94
Figure 7.5 Conflicting volume, phase 4, Northwestern and Stadium, September 19, 2007	94
Figure 7.6 Event diagrams for preemptions at US 36 and Carroll Road	95
Figure 7.7 Duration of preemption events at US 36 and Carroll Road	95
Figure 7.8 Bus performance with transit signal priority (simulation data)	96
Figure 8.1 Cumulative data and communication activity at State Route 37 and Banta Road	97

Figure 8.2 Longitudinal analysis of infrastructure and application data for entire system	97
Figure 8.3 Determining deficiencies in current communications health	98
Figure 8.4 Mapping deficiencies in the data collection infrastructure	99
Figure 8.5 Significant detector errors per 24-hour period for an arterial (SR 37 in Noblesville, Indiana)	100
Figure 9.1 Travel times: Westbound, US 36	102
Figure 9.2 US 36 Travel time CFDs for three periods during a weekday	103
Figure 9.3 June 2012 weekday travel time distributions for 11 a.m. and 5 p.m. hours	104
Figure 9.4 Bluetooth monitoring stations and subsection segment regimes on SR 37 south	105
Figure 9.5 Southbound SR 37 travel time analysis during the p.m. peak period	106
Figure 9.6 SR 37 corridor control delays using crowdsourced data	107
Figure 9.7 Travel time data from US 31 retiming in Kokomo, Indiana	108
Figure 9.8 Aggregated weekly travel times over 16 months	111
Figure 9.9 Aggregated monthly travel times over 16 months	112
Figure 9.10 Timing plan 4 (13:00–15:00) median monthly travel times	113

1. INTRODUCTION

*What gets measured gets done,
what gets measured and fed back gets done well,
what gets rewarded gets repeated.*

– John E. Jones

Traffic signal system operations is a special endeavor that falls within traffic management, yet requires a particular set of expertise and resources. The impacts of signal operations are often underestimated in the programming of budgets and staffing levels in many agencies. Even when adequate budget and staff are available, it can still be challenging to allocate resources if there is limited knowledge about which locations are operating well and which are not. Although improved technology has created a wealth of signal controller special features, sometimes with layers of responsive and adaptive control operating above, a lack of adequate performance reporting from most systems to date means the actual quality of operations is unknown for the vast majority of signal systems.

Adequate performance reporting means that enough information is provided to the system operator to know whether the quality of operations is satisfactory, as well as to suggest what aspects of control might have an impact. Currently, there is considerable emphasis on establishing transportation system performance measures; this has been underscored by dedicated language in the MAP-21 transportation bill, passed by the 112th Congress, specifying the need for better system monitoring and performance reporting. The types of information generally obtainable today are limited to equipment inventories, number of engineer and/or technician work hours invested, and other types of benchmarks that might be called “input-oriented.” Although such information has its uses, it does not directly reveal the quality of operations. Assumptions have to be made about the impact of the investment.

The purpose of this document is to compile a suite of *control-agnostic, discrete event-based* performance measures for analyzing the actual operation of traffic signal systems, applicable within any signal system and for all types of operations. The operational analysis is based on the inception and termination of protected and permitted rights-of-way, combined with simultaneous measurement of demands. From these, various types of performance information can be compiled:

- Validation of expected performance, including functionality of subsystems (e.g., detectors) and communication links.
- The efficiency of capacity utilization on lanes/movements at local intersections.
- The effectiveness of progression through coordinated lanes/movements in the system (relative success or failure of progressive timing).
- The characterization of delay and travel time.

This monograph was produced with support from a Pooled Fund Study (I) led by the Indiana Department

of Transportation (INDOT) and supported by the state transportation agencies of California, Georgia, Kansas, Minnesota, Mississippi, New Hampshire, Texas, Utah, and Wisconsin, and by the Chicago Department of Transportation. Most of the material presented herein is inspired by a series of previous studies carried out at Purdue University and the University of Minnesota in the past decade, which have been made possible under a series of national (2–4) and state-level projects (5–8). This monograph is largely a synthesis of this preceding research, including some recent additions to the growing portfolio of operational performance measures.

1.1 System Perspective: Processes and Stakeholders

A traffic signal is a component of several different systems having different potential perspectives for measuring performance. First of all, the traffic signal serves a particular purpose for which it was purchased and installed: regulation of movements at an intersection. However, this activity cannot be accomplished effectively without maintenance of equipment at the intersection, and the physical plant of a traffic signal itself comprises a system. Additionally, every intersection is part of a roadway system, which may contain multiple signalized intersections. It is often desirable to coordinate the activities of individual intersections to achieve objectives related to the overall roadway performance.

Figure 1.1 conceptualizes a hierarchy of activities involved in the enterprise of signal systems operation. This forms the basis of the development of performance measures in this monograph. Each layer in this diagram depends upon the successful implementation of all the layers beneath it.

- *Equipment procurement and installation* comprises the placement of infrastructure in the field, including all of the physical components, the power and communications infrastructure to keep them functional, field instrumentation such as detectors, and network-level elements such

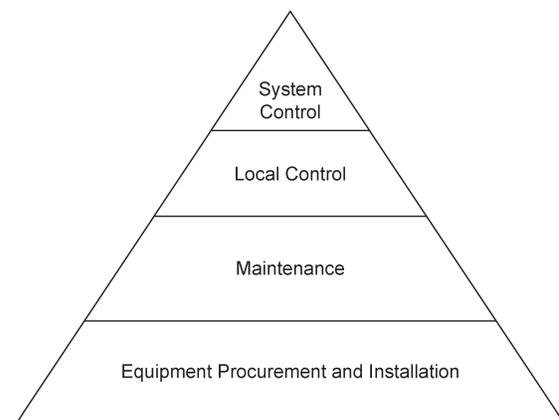


Figure 1.1 Hierarchy of needs in traffic signal system operations and maintenance.

as advanced transportation management systems (ATMS).

- *Maintenance* is the set of activities necessary to keep the system functional. A traffic signal systems is set up to keep running at a basic level, even with the failure of many of its various elements. However, the quality of service degrades as these are lost. When detectors fail, the associated phases revert to fixed-time operation. When communications fail, coordination is degraded or lost. Sometimes, these losses of functionality are difficult to perceive. Actually, the subtle nature of many of these failures makes the problem of proper signal maintenance all the more important.
- *Local Control* is the set of activities that cause signal indications to change, thereby allocating right-of-way at the intersection. In all modern, real-world signal systems, this is executed by the local controller using various switching rules. The main objective of local control is to provide efficient and equitable service to all of the various modes and movements demanding right-of-way at the intersection.
- *System Control* is the set of activities that influence the local control, coordinating the behavior of neighboring intersections. The main objective of system control is to establish coordination through a group of intersections to provide scheduled movements of traffic through a network of roadways. Good-quality system operation requires effective local control and a healthy communications infrastructure established in the lower levels.

Altogether, the entire range of activities encompassed in this vision comprise the operation of a traffic signal system. In developing performance measures, it is important to consider the organizational processes to be informed by those measures, and the stakeholders who will be making use of those measures. Figure 1.2 shows an organizational flowchart explaining the process of operating a signal system and where the opportunities lie for developing performance measures.

All agencies have some operating plan, formal or informal, that determines how signal-related activities are carried out. Measurements of invested resources present the first opportunity to develop performance

measures. For example, it is possible to track the number of engineer and/or technician hours invested, the amount of money spent on equipment upgrades, the amount of time since the last signal retiming, and other such metrics. The greater the investment, it is implied, the better the service. A recently retimed signal, or an intersection where a considerable investment has been made, is expected to have better performance. However, there is no real feedback to tell whether performance has actually improved. Although such metrics are helpful for internal review and management, this monograph does not investigate *investment-oriented* metrics.

The second opportunity for developing performance measures is to collect data on the results of signal operation, which is shown on the other side of system operation in Figure 1.2. By monitoring the individual events at intersections, it is possible to evaluate the quality of service. The analyst can determine whether the control plan is effective, whether the equipment needs maintenance, or where capital investments might be needed. This monograph focuses on this *outcome-oriented* performance measurement opportunity.

The third opportunity for performance measures is based on public reporting. The traveling public interacts with signal systems, and this sometimes generates complaint calls. For some agencies, complaint calls are the only real system feedback available. However, because such calls are created by perceived performance rather than actual performance, they are not always accurate, which can make the information challenging to use.

Figure 1.3 examines the process for signal timing design in greater detail. Six steps are defined. Although in practice these do not always occur in the order shown here, the same considerations are typically present at some level in a timing activity. Step I is to identify the objectives. Next (Step II), traffic data are obtained, such as turning movement counts, and applied to a model to design the timing plan (Step

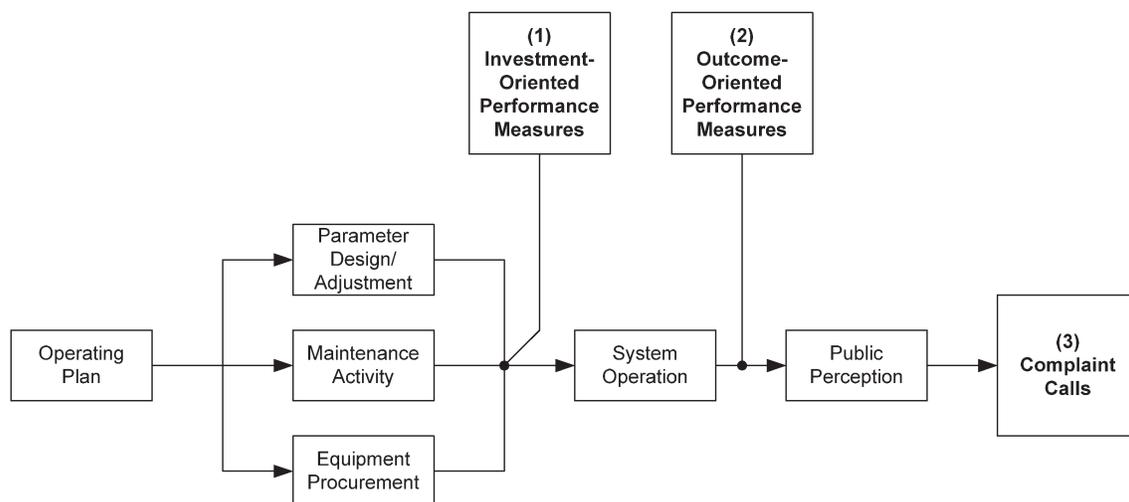


Figure 1.2 Context for signal system performance measures.

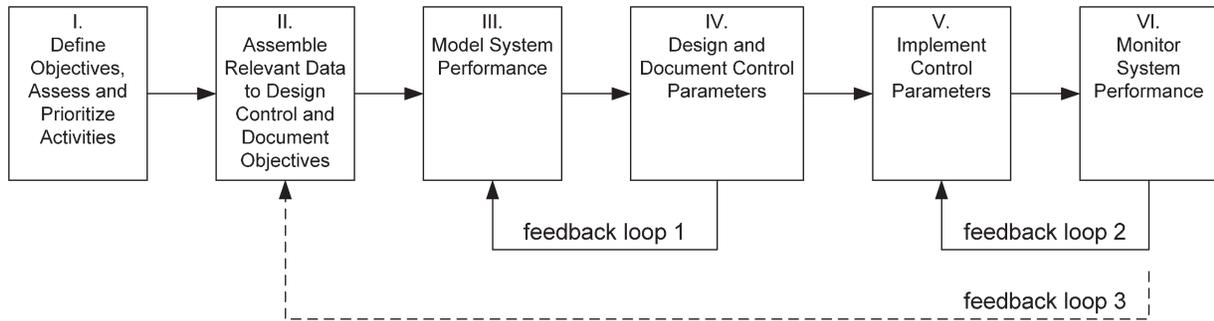


Figure 1.3 Parameter design/adjustment process.

III). The design process is often iterative, with the model results (Step IV) being tweaked in the model and tested in simulation. Next, the plans are implemented (Step V) and, ideally, monitored (Step VI) to evaluate the quality of operation.

Three feedback loops are shown in Figure 1.3:

- The first involves an interaction between the traffic engineer and the design software, which is a very data-rich process wherein the engineer shapes the timing plan based on near-immediate software feedback.
- The second loop, through which field data would measure the real performance of the timing plan, is relatively weak in many current systems. The turning movement count data collected in Step II typically only represent a portion of the day. For example, conditions for off-peak and weekend conditions are rarely known, because of the expense of obtaining them.

A third feedback loop indicating the need for more substantial action such as a complete retiming is almost completely missing in current systems. Instead, systems are retimed on an arbitrary schedule or in response to complaint calls.

A similar procedural concept can be identified for maintenance activities, as shown in Figure 1.4. Similar to the design of the control plan, the objectives are defined in Step I. The next step is to collect information about where maintenance is needed, in Step II. After this, the necessary maintenance activities would be performed (Step III), and, finally, the performance of the system would be measured (Step IV). There is an opportunity for data-rich feedback to inform the technician as to the locations where action is needed. At present, however, most systems do not provide

reporting tools to search for detector errors or other heuristic measures to indicate where problems exist. Without adequate information, maintenance is carried out on an arbitrary schedule, where technicians are sent regularly to intersections regardless of whether they have a problem or not, and where problems that do occur must wait until the next scheduled maintenance time for resolution.

Last, the equipment procurement process itself can be broken into a series of steps as in Figure 1.5. Starting from the agency objectives (Step I), the equipment needs are generated (Step II), after which the agency begins seeking and evaluating proposals (Steps III–IV). After selecting the equipment, it is installed (Step V) and then monitored (Step VI). Similar to the design of control parameters, the proposal evaluation process involves considerable feedback (loop 1), as the tradeoffs are considered between potential investments. Once a system is installed, there is usually a period of monitoring to ensure that the new equipment is functional (feedback loop 2). Ideally, there would be a third feedback loop in which information from the system would assist determinations about equipment needs in Step II (feedback loop 3). This is not very common in practice, because the infrastructure often does not support it. Instead, resources are allocated according to facility age, the time elapsed since the last site visit, or engineering judgment.

This monograph proposes and demonstrate a suite of performance measures for improving these three processes by allowing different aspects of system performance to be documented and tracked over time. This includes a series of operational performance

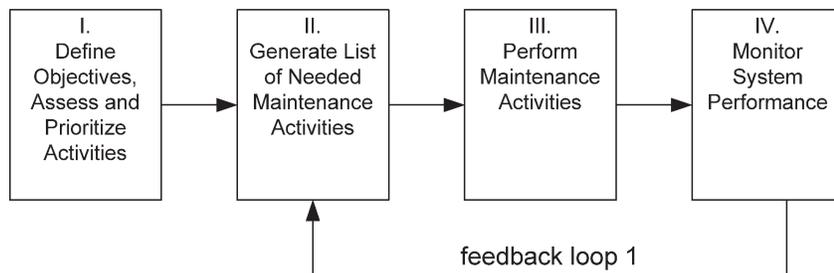


Figure 1.4 Maintenance process.

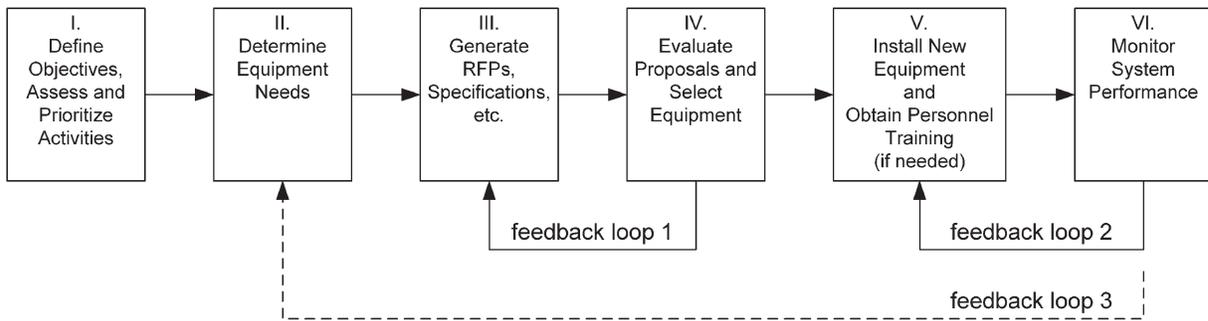


Figure 1.5 Equipment procurement process.

measures broken down by local control and system control objects, to assist the signal timing process (Figure 1.3), together with a series of maintenance-oriented performance measures (Figure 1.4). Both of these would contain valuable information for those engineers in the process of determining where new capital equipment should be deployed in the system (Figure 1.5).

1.2 Physical Elements of a Traffic Signal System

Figure 1.6 shows a conceptual view of elements in a traffic signal system, from the cabinet at a local intersection to a system-level advanced traffic management system (ATMS). A collection of subsystems exist within the cabinet. The signal controller is the primary component, as it determines the various signal indications. In Figure 1.6, this is labeled as the “primary” controller. Some traffic control systems have “secondary” layers of control implemented by the ATMS or by secondary controllers (such as “master” controllers). One common system-level function is to synchronize local controller clocks to maintain conventional coordination. Other strategies combine local-control routines processed in the primary controller with system-control routines executing in a secondary controller.

As technology has advanced, controllers have gradually evolved from single-purpose devices into computing machines capable of running various applications. For example, the type of data logger collecting the traffic information presented in this methodology executes on the controller in parallel with the software running the intersection. This trend will likely accelerate in the future. The “Advanced Transportation Controller” (ATC) specification incorporates this concept (9). More recently, an IDEA project sponsored by FHWA has led to the development of an open-source controller that can be executed on any Linux machine (10). Australian researchers, meanwhile, are working on a next-generation controller that will support “plug-in” traffic control software (11). These types of developments may lead to future controllers that are platforms for running applications, whereby new control strategies are implemented with software rather than hardware. From this perspective,

it seems likely that future controllers will, perhaps universally, feature advanced data-logging capabilities; these data will become more important as the sophistication of the control systems increases.

In addition to the primary controller, the *malfunction management unit* (MMU) is a sort of local secondary controller that monitors the signal state, and can take over the intersection (by putting it into flash) when the primary controller produces output that violates certain rules—such as the display of conflicting greens or a yellow time that is too brief.

Some kind of access point for external communication is available in most modern signal cabinets, which has traditionally supported “closed-loop” control. Although this terminology implies that a *control* feedback loop exists, in practice this has mainly provided a clock synchronization function. The low bandwidth of dial-up connections has historically constrained the development of greater network functionality. Some manufacturers have provided both the capability of monitoring split times and detector occupancy. These data are usually aggregated in the controller so that only small amounts of information need to be passed through the low-bandwidth connection.

Of course, much *has* been done to provide for “area traffic control” over the years. As early as the 1930s, basic traffic responsive systems were developed in the United Kingdom using analog electronics (12). In the late 1950s, a number of cities, such as San Jose and Toronto, began investing in computer systems to schedule and select timing plans. Commercial traffic-responsive systems appeared around the same time. The first adaptive control systems appeared in the late 1970s (13). These systems have evolved substantially since then (14), and there are many such systems that an agency might today consider acquiring to provide advanced functionality. Such an acquisition would be a capital investment, and the scope of the solution would be limited to a corridor or a street network, rather than a region or entire agency.

Recent developments have begun to change the state of communications. The proliferation of wireless devices, and the reduction in their costs, has begun to affect the traffic industry. It is now possible to establish IP communications in geographically distributed regions at a relatively low cost (15). With the

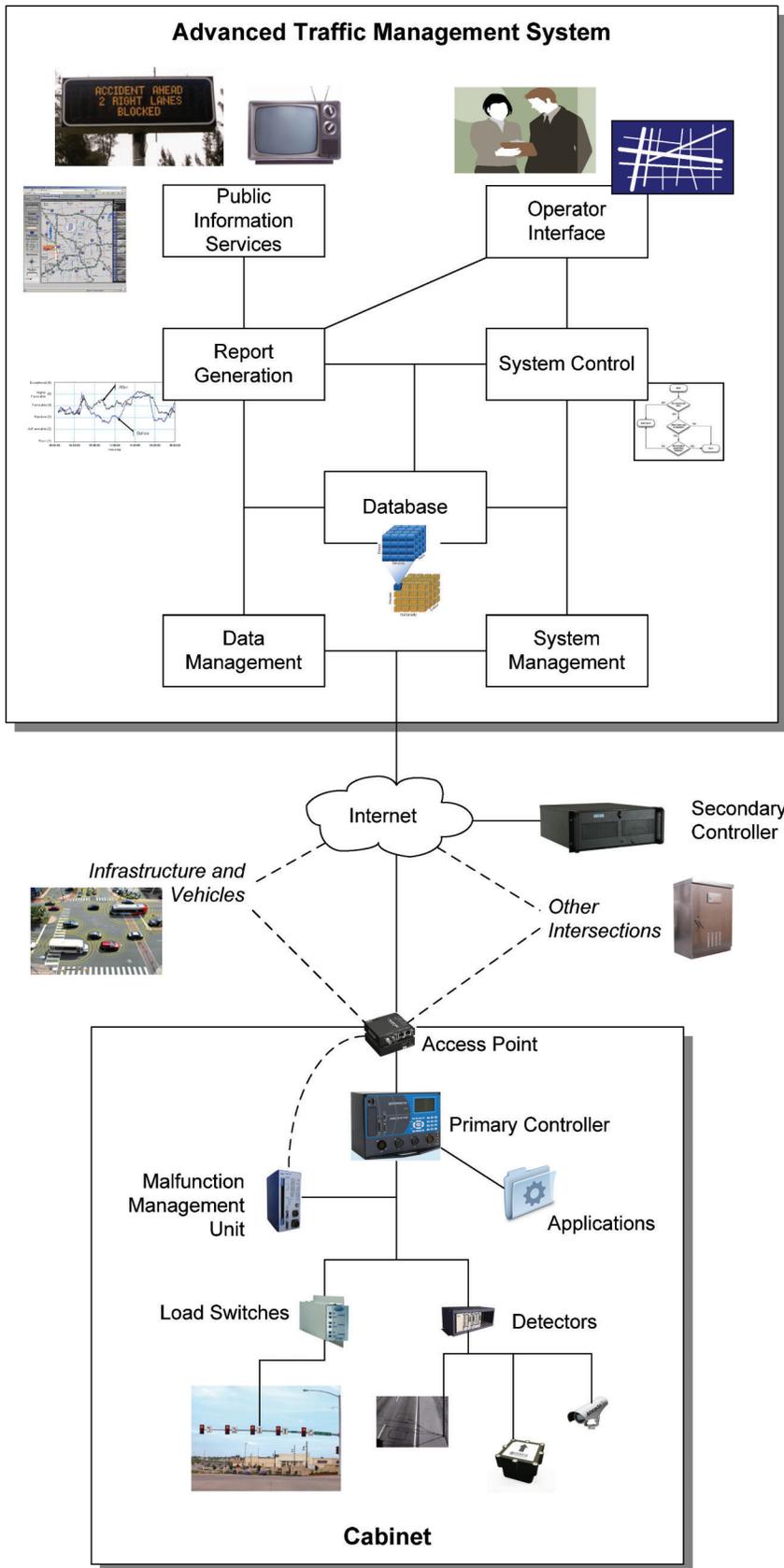


Figure 1.6 Overview of elements in a traffic signal system.

improvement of communications to the field, it becomes possible to obtain rich data sets from the field, providing accurate logs of controller activity and vehicle behavior from which performance measures can be extracted.

System monitoring is vitally important to assess and maintain the quality of service in a signal system. An ATMS could serve as a platform to execute system monitoring functions, either as a server-driven “pulling” of data or as a client-driven “pushing” of data. Within such a system, several components can be identified with particular ATMS functionalities (Figure 1.6):

- A *system control engine* of some sort perhaps best reflects the traditional core functionality of any ATMS: the centralized management of control—a family of functions ranging from the synchronization of controller clocks to the issuing of commands influencing local controller timing and scheduling.
- Within the ATMS, there are of course the *control processes* executed by the system control engine, which are controlled by various parameters (stored in a database) and programmed by the user.
- A *data collection engine* is needed to collect data from local intersections to develop performance measures.
- At least one *database* of some kind archives and manages the data reported by intersections, as well as the ATMS itself in the execution of its various processes.
- A *reporting engine* is needed to process raw data to construct performance measures, and to assemble this information into coherent and digestible reports for a variety of end use cases.
- An *operator interface* enables the system user to monitor and program various ATMS control processes, as well as to observe traffic and view performance reports. The ATMS also provide a means of programming subsystem elements such as local controller timing plans.
- In some cases, the ATMS would support *public information services*, either directly through website and/or mobile device applications or indirectly by supporting reports in the media.

Future work will likely continue to improve communications as well as the type of information that can be collected. Of particular importance in this area is the development of “Connected Vehicle” technology, which would allow direct communication between vehicles and roadside infrastructure (such as signal controllers), and among vehicles through vehicle ad hoc networks (VANETs). In other words, the type of information today collected by detection would be vastly extended. Current research emphasizes collision avoidance and infrastructure-to-vehicle communication (such as the amount of green time remaining for a particular phase or what time a light is expected to change). Even before such technology comes to market, some developers of smartphone applications have created driver advisory systems that attempt to predict phase changing times to provide recommendations as to whether vehicle speed should be maintained or changed (16,17). The possibilities of cooperative intersection control might even challenge the notion that advanced intersection control must be infrastructure-driven. In fact, if vehicles can

effectively create their own “green waves” by dynamically adjusting speeds and routes, it might be more advantageous for intersections to operate more predictably.

This monograph focuses on the details of creating performance measure views relevant to deployment within the reporting engine. Its purpose is twofold. One is to document a variety of metrics that have arisen over the course of several years of continuous research, as well as lessons learned during numerous pilot deployments. The second purpose is to serve as a reference for investigating and interpreting these performance measures and to assist agency personnel in making the best use of the data and of their traffic signal infrastructure.

2. INFRASTRUCTURE REQUIREMENTS

This chapter discusses the necessary system elements needed to implement the performance measure methodology. This includes a discussion of vehicle detection, data acquisition equipment, and communication infrastructure.

2.1 Vehicle Detectors

A key requirement of the performance measurement methodology is an ability to measure the number of vehicles utilizing the intersection. Outside of central business district areas, most intersections feature some type of detection to actuate phases. Historically, many of these detection capabilities have been fairly limited. To cut costs, for example, some agencies do not provide upstream detection, and in some cases there is no detection whatsoever on the “main street,” with the rationale being that the signal timing will be set up to always give the maximum amount of time to those movements. However, as detection technologies continue to improve, it is becoming less difficult to add new detection zones, as in the case of minimally “invasive” systems such as video or infrared detection or wireless magnetometers. In any case, an intersection that lacks detection on any particular movement would require the installation of new detection before *any* substantive operational improvements can be expected (other than by conventional retiming, with the necessary manual data collection).

The methodology presented in this monograph is compatible with existing detection plans at many intersections. It is desirable to have detection on every lane at the intersection (having two adjacent lanes on one channel is acceptable although less accurate). To evaluate signal coordination, setback detectors are needed to measure vehicle arrivals. For capacity performance measures to evaluate local control, either stop bar or advance detectors will suffice.

A distinction must be made between presence and count detector output, particularly if the reader has not had time to meditate on these differences before. In signal operations, vehicle detectors currently provide a “contact-closure” binary output, comprising either ON

or OFF (DETECT or UNDETECT) states. Many detectors can be set up to provide one of two types of output: *presence* detection that stays ON for as long as any vehicle is in the detection zone and *pulse* detection that sends a very short ON signal upon the entry of the vehicle into the zone. Typical detection zones range approximately from 30 to 50 ft in length. Because it is not always certain where vehicles will actually stop in relation to the stop bar, rather long zones are typically needed to cover a flexible range of pavement to accommodate the situation where a single vehicle arrives at the intersection. Four 6-ft loops with 9-ft spacing (loop edge-to-loop edge) would comprise a 51-ft zone.

Figure 2.1 illustrates the difference between pulse and presence data for a typical stop bar detection zone. This example assumes that the detector provides both presence and vehicle counts through detailed analysis of the detector response. These are received by the controller on two separate detector channels.

- At time t_0 , there are no vehicles in the detection zone, so the state is OFF (0).
- At t_1 , the first vehicle enters the detection zone and the presence channel turns ON (1), while a short pulse is registered in the count channel.
- By t_2 , the first vehicle has moved to the middle of the detection zone and the presence channel is still ON (1), while the count channel is OFF (0).
- At t_3 , the first vehicle has very nearly exited the detection zone, but the front of a second vehicle has just entered the zone behind it.

The count detector is able to see the influence of the second vehicle by analyzing the analog data (e.g., the change of frequency in the inductive loop change response), and it registers a second pulse on the count channel. However, there is no evidence of the second vehicle in the presence channel. Only when the detection zone is completely empty does the presence trace fall to OFF. Although the presence channel does provide vehicle occupancy, the count channel is needed to measure the number of vehicles.

Figure 2.2 shows a map of potential detector locations on a typical signal approach.

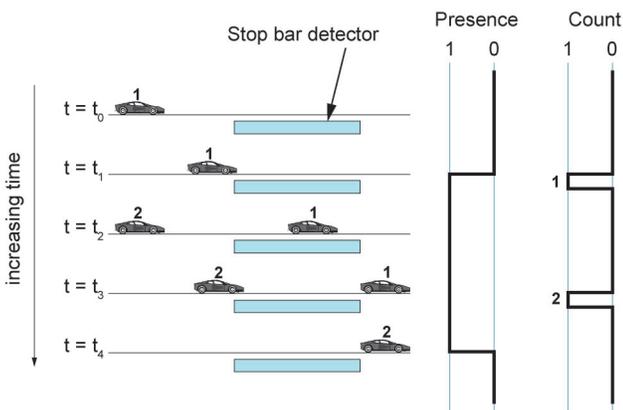


Figure 2.1 Presence versus count (pulse) detection.

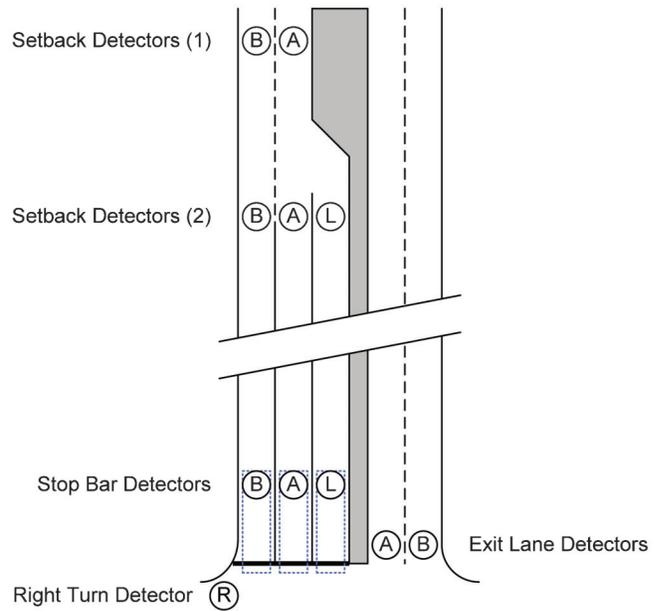


Figure 2.2 Potential detector locations on a typical signal approach.

- *Stop bar detectors* are often able to obtain accurate counts on each movement at an intersection, with some caveats. In shared lanes, it is unclear to which movement a given count belongs, and some assumptions must be made if movement-specific counts are desired. The detection range of loop detection zones usually find counts for vehicles *entering* the detection zone, and the processing of the count data may introduce up to 1 second of latency. While this is acceptable for vehicle counts for a cycle length or a green interval, the resolution of the count accuracy is not sufficient to analyze vehicle headways.
- *Exit lane detectors* are less frequently seen in practice, but can also establish vehicle counts. Because vehicles typically do not form queues over these detectors (except in gridlocked conditions), headways can be measured directly from detector presence (18). Depending on the intersection geometry, it may be challenging to accurately assign counts to movements because of right turns on red.
- *Right-turn detectors* provide detection on extended right-turn lanes, where vehicles can sometimes stop well ahead of the stop bar while waiting for gaps. Many signals operated by INDOT feature such detectors. This can assist in developing accurate turning movement counts, particularly for shared lanes.
- *Setback detectors* provide “dilemma zone” protection by reducing the likelihood of phase terminations occurring when vehicles are too close to the intersection to be able to safely stop. Such detectors are essential for evaluating coordination, by making it possible to relate vehicle arrival times to the local signal state. Depending on the geometry of the intersection, these detectors might exist before or after the addition of turning lanes. In Figure 2.2, the first group of setback detectors would count both through and left-turning vehicles without being able to distinguish between the two, while the second group provides more accuracy by having the left-turn lane separated, although some lane changing

behavior may still occur past the detection zone. In the former case, where the detectors are situated before turning lanes begin, then the volume of the turns should be subtracted from the setback detector counts to obtain a count for the through movement.

Although it has historically been challenging to obtain automatic vehicle counts from vehicle detectors, the quality of data has improved considerably in the past decade. High-quality vehicle counts have been obtained from inductive loops (19,20) as well as video detection systems (21). For signal timing, INDOT has largely replaced manual counts with automatic counts. In the preparation of this monograph, it was decided to compare automatic and manual to characterize the effectiveness of automatic counting.

The intersection selected for this study was US 36 (Pendleton Pike) and 56th Street in Indianapolis, Indiana. Figure 2.3 shows an overview of the intersection. There are stop bar detectors on 56th Street and setback detectors on Pendleton Pike. Four hours of video were recorded at the intersection during fair weather conditions on May 17, 2012, while counts were automatically collected at the same time. Each lane at the intersection had a separate detector channel, with the exception of the southbound right turn. Figure 2.4 shows an example of the video recording. Northbound, southbound, and westbound movements were distinguishable in the video.

Figure 2.5 shows the cumulative cycle-by-cycle counts for six lanes at the intersection. The automatic and manual counts have good agreement in most cases. The westbound lanes (Figure 2.5a, Figure 2.5b) track each other well, with a few minor divergences of the two traces. At the stop bar, the count pulse is obtained by post-processing of the inductance state, which leads to occasional miscounts. Automatic counts from the northbound detectors (Figure 2.5c, Figure 2.5d) also

exhibit a high degree of agreement with the manual count. Because these are setback detectors, the presence is used for automatic counting directly, without post-processing. The outside lane has a very slightly higher automatic count, which is likely due to the inclusion of a few right-turning vehicles in the stream.

The southbound lanes (Figure 2.5e, Figure 2.5f) show a much higher automatic count than manual count. This is because all of the right-turning vehicles are in the same traffic stream as the through vehicles, whereas the manual count included only through vehicles. Figure 2.3 shows that the right-turn lane begins *after* the setback detectors, so traffic over the setback detectors includes all right turn vehicles. Over half of the vehicles passing the setback detector in the outside lane are likely right turns (Figure 2.5f), as well as a smaller portion of vehicles in the left lane (Figure 2.5e). The southbound right-turn lane was not visible in the video, so a total approach count was not possible in this case.

The methodology requires **vehicle counts** to be available at the intersection. It is desirable for each lane to have an individual detector, but acceptable performance measures can be obtained with parallel lanes joined under a single detector and at intersections that lack detectors for noncritical permitted movements, such as right turns.

2.2 Data Acquisition

After detection, the next critical piece of field equipment needed to implement the performance measure methodology is a means of recording the event data. There are several ways to accomplish this using current technology.

- As part of collaborative research efforts between INDOT, Purdue University, and several controller

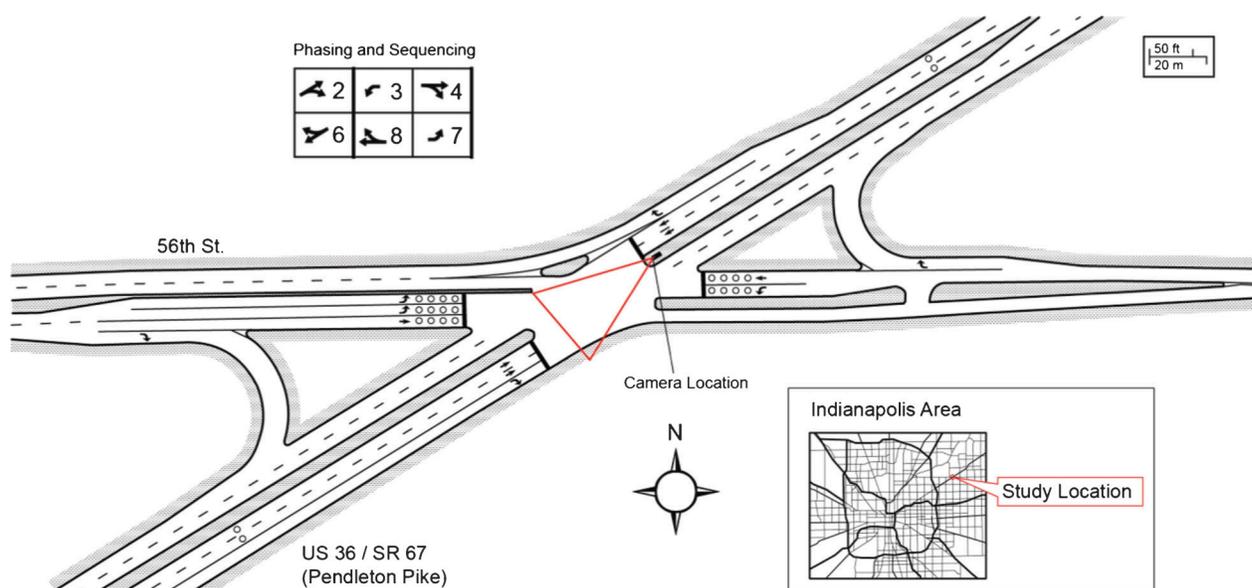


Figure 2.3 Pendleton Pike and 56th Street, Indianapolis, Indiana.



Figure 2.4 Example frame from video for visual counting.

manufacturers (22), there are now several¹ commercially available controllers with onboard data collection software. The specifications for the data encoding have been published (23) and could be implemented by any manufacturer. Data acquisition consists of automated FTP downloads from each controller.

- Researchers at the University of Minnesota have developed an industrial PC that is placed in the signal cabinet and reads the signal status by monitoring the SDLC bus (8).
- Recently, researchers at the Texas Transportation Institute reported on the development of a portable platform for signal status monitoring (24).
- The Performance Measurement System (PeMS) developed at the University of California at Berkeley and operated by Berkeley Transportation Systems (which was later acquired by Iteris) has introduced signal status monitoring (25).
- A system called Trafmate, sold by TrafInfo Communications, Inc., is capable of automatically downloading files from various detector systems within a controller cabinet, and can log green times and cycle-by-cycle counts (26).
- Some advanced control systems and “central system” software can display real-time signal status. The status itself contains much of the information needed for signal performance measurement. It is not difficult to envision the storage of that data in a relational database for later querying. Historically, this was not a practical use case, given the communications and computing infrastructure then available. Today, these limitations have eased, and high-resolution signal event logging may emerge as a common application of these types of systems.

2.3 Communications

After the events from signal controllers have been logged, the data must be transported to a management center for archival and generation of performance

measures. In the INDOT system, data are harvested via a communication chain starting from the signal cabinet to a server in the central office. The data are typically stored in the signal controller or on a separate computer with Ethernet access. Most of the current generation of traffic signal controllers feature Ethernet ports that allow connection to a local router inside the cabinet. The router acts as a gateway to the Internet, or to other signal cabinets within a local area network.

Figure 2.6 shows an overview of this scheme. For agencies that cover a wide geographic area, it is not always cost-effective to implement a system-wide dedicated communication infrastructure. INDOT uses commercial cellular devices to overcome this challenge. Cellular modems are placed in enclosures with weatherproof protrusions from the tops of cabinets, allowing Internet connections to be made through the cellular network. Locally, the connection varies, as shown in Figure 2.6. In Zone 1, for example, there is no field connection, and each cabinet has its own modem. Cabinets in Zone 2 route IP communications via a local fiber connection to a master cabinet containing a single cellular access point.

Unlike data transported across private networks on dedicated infrastructure, communication through commercial networks is generally visible to the public. To address this issue, a measure of security can be achieved by enabling virtual private networking (VPN) features on the access point routers and in the management center’s router. This method encrypts the data using private keys before transporting the information over the network. Although no communications carried over the Internet can be perfectly secure, including VPN, it has worked well in practice in Indiana.

In situations where cellular connections are not feasible, or where communications issues are persistent, an alternative may be to use a local data collection device in the cabinet, with scheduled retrieval of the data by temporary connections or by physical visits to the site. The authors of this document recently set up an automatic data downloader running as a scheduled task in Linux on a compact computer² at a total cost of under \$50 (Figure 2.7). This has been deployed at six locations on US 231 in West Lafayette, Indiana, in place of cellular modems. The data can be retrieved on weekly, biweekly, or even longer intervals, as the computer is theoretically capable of storing many years of data.

At the management center, application services can be executed on servers connected to the VPN to ingest and process the data (For any intersections not on the network, the data can also be manually ingested). The data management process typically follows the following steps: data download, format normalization, and data archival.

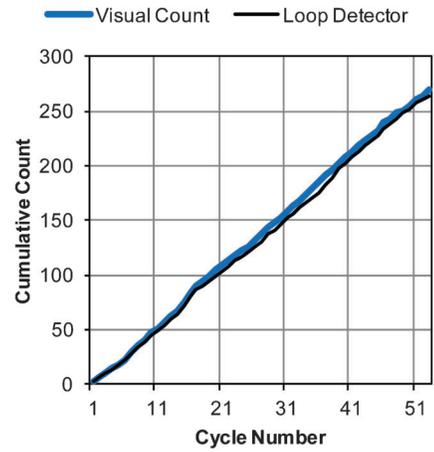
- *Data download.* A management center server opens connections to each controller and downloads event data. This can be scheduled to run at set intervals or triggered manually by an operator. The length of the scheduled download intervals varies according to communication

¹At the beginning of 2014, these include the Econolite ASC/3, Peek ATC, Siemens SEPAC, and Intelight NEMA X2, and a Naztec controller in development.

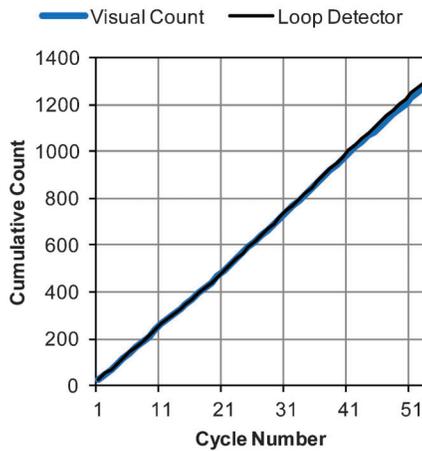
²<http://www.raspberrypi.org>



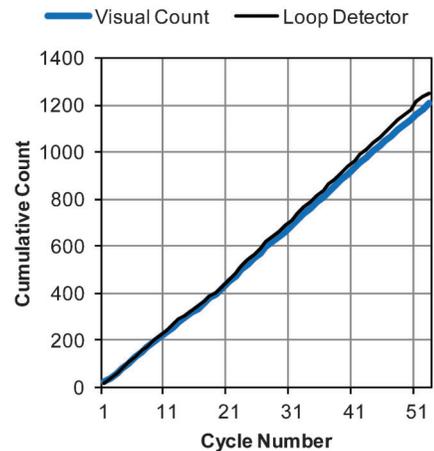
(a) Westbound left (stop bar)



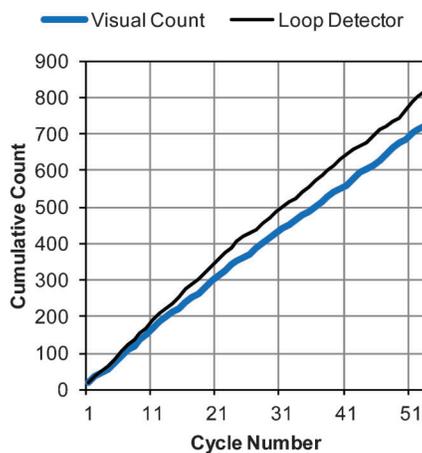
(b) Westbound thru (stop bar)



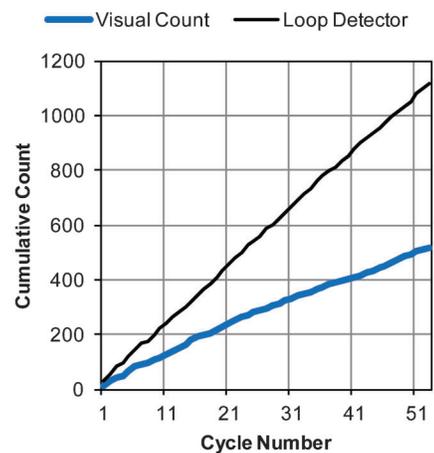
(c) Northbound inside lane (setback)



(d) Northbound outside lane (setback)



(e) Southbound inside lane (setback)



(f) Southbound outside lane (setback)

Figure 2.5 Comparison of visual counts versus loop detector counts.

quality and bandwidth. The process can also be run either sequentially per controller or in parallel across many controllers, contingent on operating strategy.

- *Format normalization.* Different event logging platforms, whether they be signal controllers or computers, can

output data in different formats. Some data are stored in binary format whereas others may be compressed or in plain text. The organization structure of each event can also vary between platforms. The goal of format normalization is to convert and mold the data into a

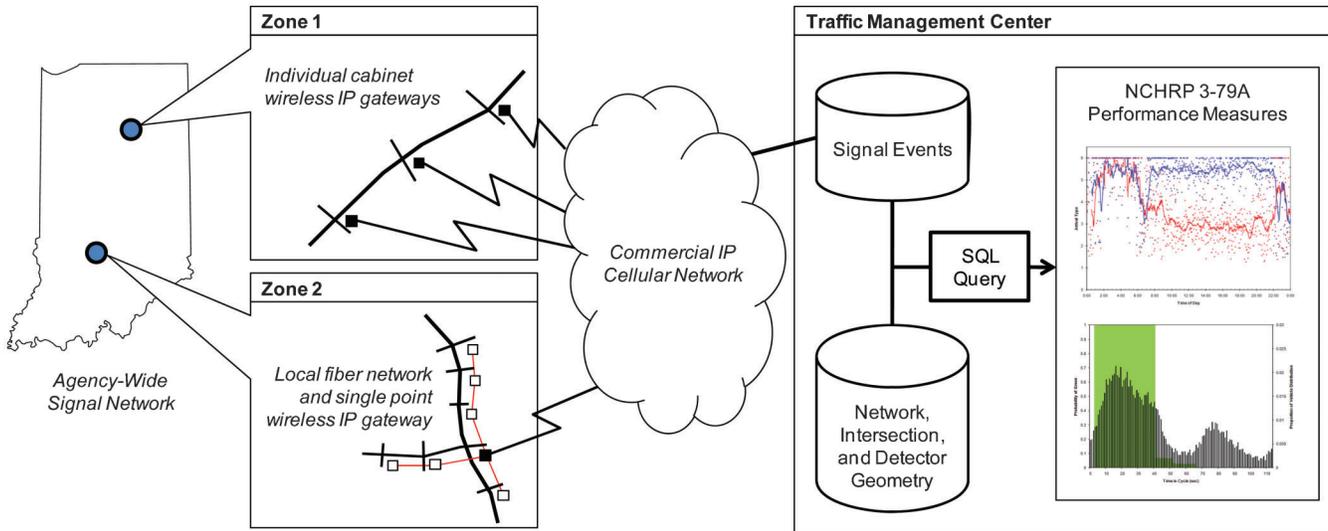


Figure 2.6 Communications architecture (15).

uniform structure so that performance measures can be generated consistently irrespective of the dynamism of event data sources.

- *Data archival.* The data are moved to a long-term storage solution. Some platforms that store the event data include relational database management systems (RDBMS) such as Microsoft SQL Server and PostgreSQL, structured file systems, and compressed archives. Performance measure generation can benefit from an RDBMS implementation because of the ease of data access and filtering as well as powerful features such as indexing and user access control. In the INDOT system, a RDBMS generates performance measure generation and archives data, and a file system structure stores the original controller files.

2.4 Data Infrastructure

Figure 2.8 shows a basic relational database schema that stores signal systems, route, detector, and event information. Each database object contains a unique

identifier field or relates to another object by a foreign identifier field. For example, an agency may keep track of a few systems in their signal network in the “system” table. This table contains a unique identifier field called “id.” In addition, there may be a number of signals stored in the “signal” table, all of which have their own unique identifier fields. The signals can be then grouped to a system by having a common value in the “system_id” field that refers back to the system record. Below are descriptions of the essential categories of signal systems performance measure data.

- *Signal and system.* These contain organizational, spatial, and Internet Protocol (IP) address information about each signal controller and its associated system. This allows pinpointing of each signal controller for data retrieval.
- *Detector.* This stores phase, presence and count channels, and physical location with respect to the intersection (direction of approach, lane, etc.).

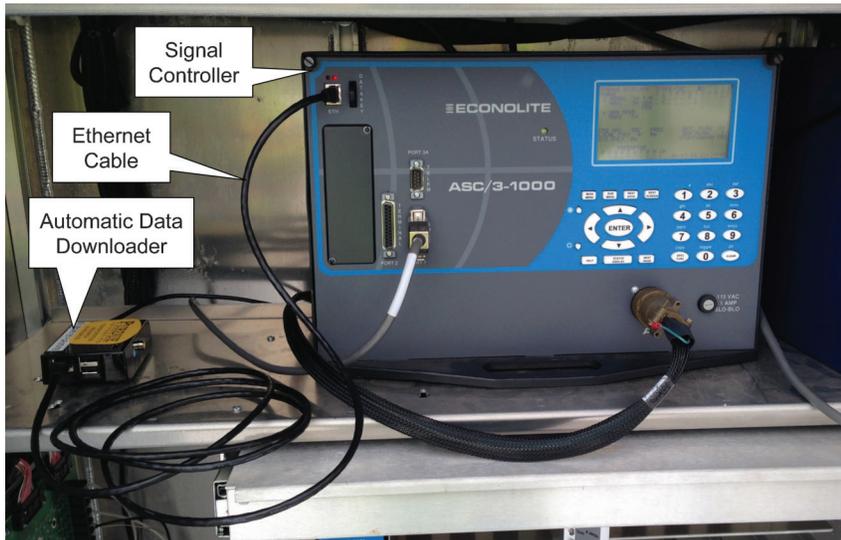


Figure 2.7 Deployment of automatic data downloader installed on a compact Linux computer.

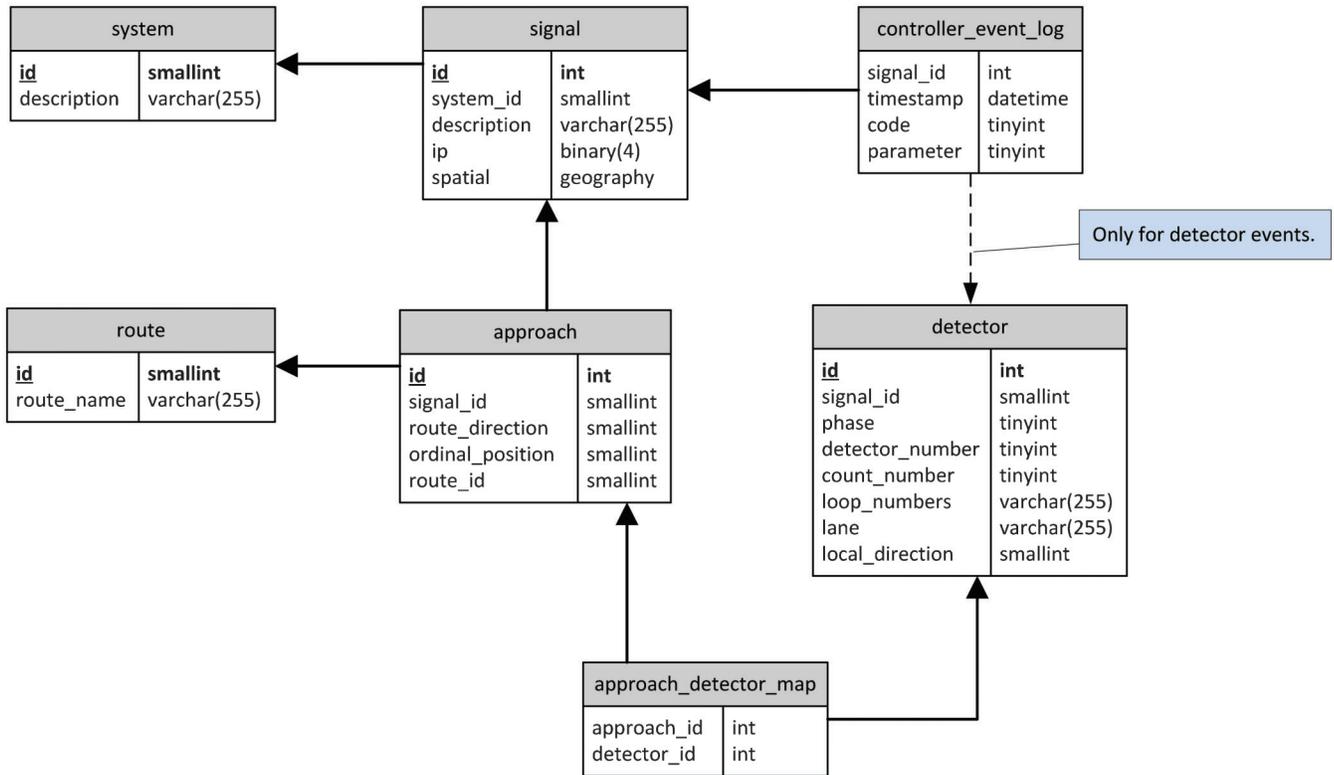


Figure 2.8 Data infrastructure (simplified example).

- *Routing and approach.* These define the ordinal sequence of signals and approach directions based along a roadway path. Detectors are mapped to approach directions to relate the detector event information to a particular approach. The routing data are needed for optimization of signal timing.
- *Controller events.* This stores the timestamps of controller events. Each record references a unique signal identifier in a multisignal database.

Because of the large number of controller events generated by signal controllers in the field, managing the archived data can be a challenge. An Econolite ASC/3 controller at a typical high-volume, eight-phase intersection with detection on all phases (across 12 or more lanes) can generate several hundred thousand events per day. As data accumulate over time, the dataset will require more time to query as a whole, thereby slowing performance measure generation. This challenge can be addressed by a combination of four data management strategies: minimizing the number of columns, using small column data types, indexing, and table partitioning.

- *Minimizing the number of columns.* When event data are stored, each record should only provide information necessary for knowing what happened at the intersection. This typically includes (1) the signal identifier, (2) timestamp, (3) event code, and (4) event parameter.
- *Using small column data types.* Data size can be reduced dramatically by using the smallest possible data type sufficient to represent each column. For example, in the Econolite ASC/3, both the event code and event

parameter values range from 0 to 255, and can be represented using one byte each.

- *Indexing.* One column of the event table can be designated as an index to speed up queries that filter on those columns, e.g., conditionals specified by the “where” clause in an SQL query.
- *Horizontal table partitioning.* It is possible to divide groups of rows in the events table into different tables. Horizontal table partitioning increases the efficiency of queries by limiting the search ranges to smaller sets of data. Figure 2.9 illustrates a sample events table partitioned by the timestamp column. The data range of each partition spans a one-month period.

Although it is not a use case that has been pursued by the authors, it is possible that a subset of data could be logged to further economize on data storage requirements. Appendix A provides a list of performance measures described in this monograph, and the minimal set of events needed to produce them. Another possibility could be a distributed model where processing of the performance measures is accomplished on the controller, and only the aggregated performance measures are logged in the central system. The authors believe it is always better to preserve the raw data when possible; however, for some use cases, a data subset or local aggregation approach may be appropriate.

2.5 Data Size

Depending on the size of the signal network and how many event records are retained over time, the amount

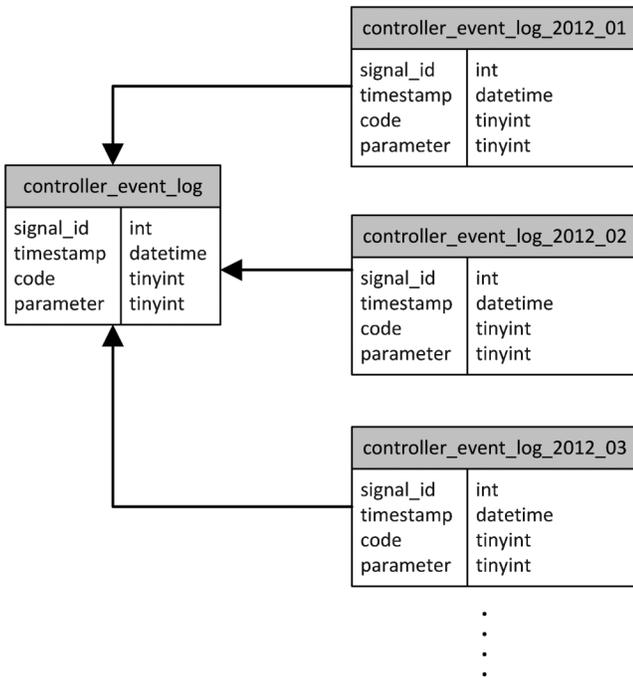


Figure 2.9 Horizontal table partitioning.

of disk space necessary to store the data may vary. Each event logged within a signal controller event file takes up four bytes of disk memory (timestamp, event code, and event parameter). For the same event to be transferred into a relational storage database structure, additional storage is required to identify the signal associated with the record, as well as a longer six- or eight-byte timestamp that includes date information. Typical timestamps in event log files only contain time information in number of seconds elapsed relative to an initial time stamped in the file header.

Considering an event table structure using three two-byte data columns and one six-byte data column, the

storage requirement per event is 12 bytes. For a single high-volume intersection of about 40,000 vehicles per average weekday, this takes up 468.75 kilobytes of storage daily for count detection events alone. Figure 2.10 illustrates the daily and cumulative data recorded over time for a particular high-volume intersection over a six-month period. Note that there is a weekly trend here—the weekdays tend to have higher volumes than the weekends. A loss of communications in mid-October accounts for the gap in the data at that time. This particular intersection requires about 220 MB of storage over the five-month period.

2.6 Performance Measure Implementations

To close this discussion of the data requirements, we present three example deployments of performance measure systems based on the methodology presented in this monograph.

- *Indiana Department of Transportation* (Figure 2.11). The initial deployment of the controller firmware data collector (22) was at the intersection of SR 37 and SR 32 in Noblesville, Indiana. In 2008–2009, this was expanded to four intersections (and later eight) along SR 37. At this time, a pilot system for running automatic data downloads was established, and the controller data were integrated into a server residing at the INDOT Traffic Management Center. Work began to develop a web front-end to enable the viewing of the performance measures. As of 2014, this system had been expanded to 158 intersections. System growth has scaled according to the addition of new nodes in the communications network and incremental change-out of old controllers.
- *Elkhart County* (Figure 2.12). In 2010, Elkhart County, Indiana developed the first specification for an advanced traffic management system (ATMS) with specific language to require a performance measure capability. The result was the procurement of an Econolite Centrac system with the capability of automatically downloading the high-resolution data and displaying a set of

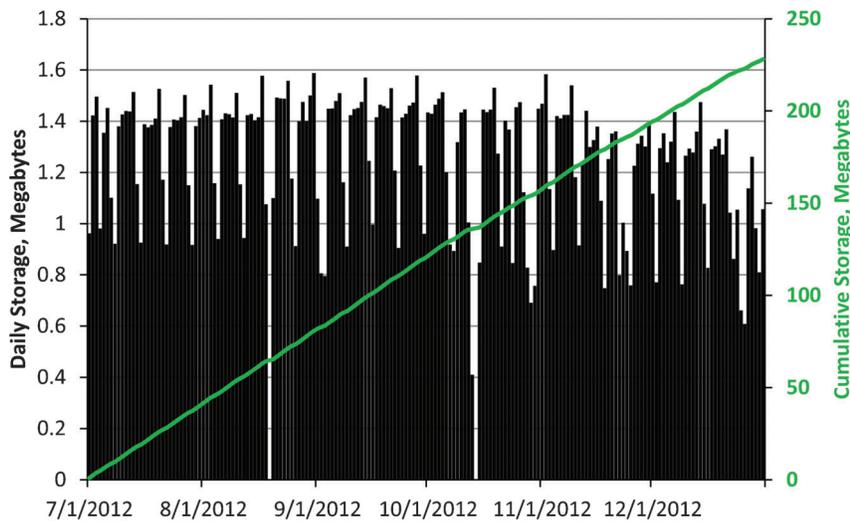


Figure 2.10 Data storage requirements over time for one intersection.



Figure 2.11 Indiana Department of Transportation performance measures website.



Figure 2.12 Elkhart County Centrac system.

performance measures. In the same year, the city of Lafayette, Indiana, procured another such system using a similar specification.

- *Utah Department of Transportation* (Figure 2.13). In the past several years, the Utah Department of Transportation (UDOT) began investing resources into the development of

performance measures based on the high-resolution event data. This led to the independent development of a performance measure website, with some similarities to the INDOT system. Because of strong prior investments in communications infrastructure by UDOT, this system has rapidly expanded to include 1,027 intersections in total in

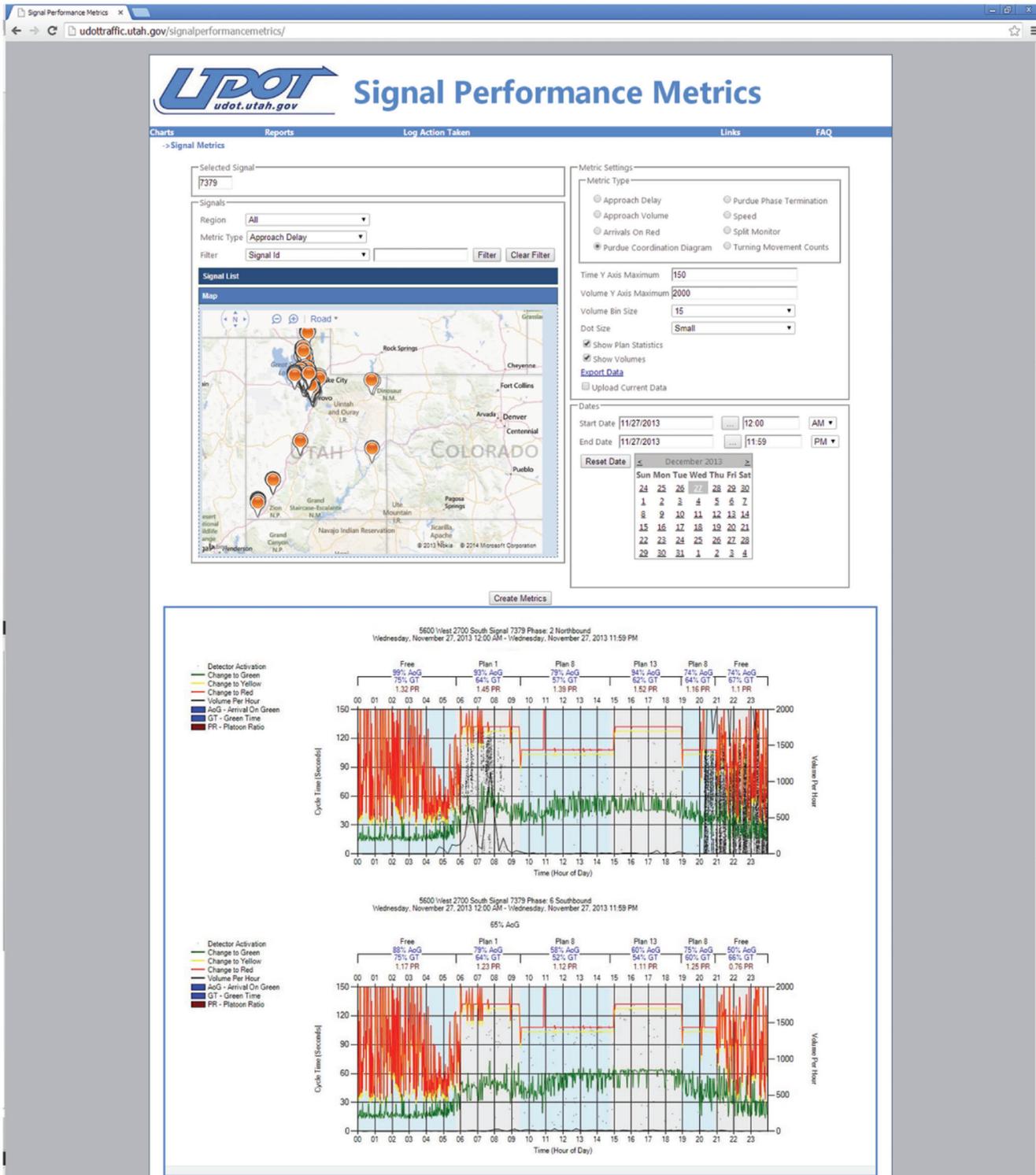


Figure 2.13 Utah Department of Transportation performance measures website.

2013. The Utah system is a good example of how such a system can be quickly deployed when there is an excellent data infrastructure in place.

3. ASPECTS OF SIGNAL TIMING

This chapter introduces basic signal timing concepts and their impacts on the operation of real-world controllers. The purpose of this discussion is to inform the subsequent development of performance measures, while providing a light background for readers less familiar with the subject.

3.1 Types of Signal Operation

Table 3.1 presents a listing of different modes of signal operation. These are organized by the type of detector data needed for local control and the type of cycling behavior applicable to system control.

- *Fixed-time* control serves each movement in a rigid sequence of green times without any detection. This type of operation is common in central business districts and similar locations, and is usually coordinated. These types of systems operate with a fixed cycle length.
- *Semi-actuated* control typically features actuated minor movements, whereas the major movements may not have detection. The major phases are recalled during every cycle, and tend to receive the greatest share of the green time. In cases where there is low demand for the major street, this type of control tends to cause unnecessary delay for side streets. This type of control can be found in both coordinated systems, which run on fixed cycle lengths, and in isolated locations that are not coordinated.
- *Fully actuated* control features detection on every movement at the intersection, including the major street. Every movement has the opportunity to be called and terminated by detector activity. Coordinated control takes place with a fixed cycle length, which can have small variations due to actuation. Isolated fully actuated intersections usually do not have a fixed cycle length, and each green begins and ends according to measured demand. It is common to place recalls to make the controller rest in green along the major street. Variable-cycle fully actuated systems are generally not coordinated, although there has been research into ways of coordinating neighboring intersections by adding extended features to the usual set of fully actuated control parameters (27).
- The fourth category of control in Table 3.1 represents control schemes featuring a *prediction* of traffic patterns. Many adaptive control systems exist within this space. Some use a fixed background cycle length to align green

times at adjacent intersections, and others use a variable cycle length. The implementation of advanced control varies considerably by system. Regardless of the underlying algorithms, however, the control decisions are always translated into the states of the signal indications, and their record of activity is seen in the trace of controller status that is their outcome.

3.2 Control Elements

Each movement at a traffic signal is governed by a particular indication (green/yellow/red for vehicles and walk/don't walk for pedestrians) controlled by a load switch in the cabinet. The load switch state is in turn controlled by a controller output channel. The maximum number of load switches is determined by the cabinet type.

There are two widely used timing systems for organizing signal output channels:

- *Interval* timing focuses on the electrical outputs of each output channel, called a *signal group*. The signal groups are scheduled in a cyclical timing program that specifies each when the signal output changes indication, and what its indication should be. Pedestrian indications are controlled by their own signal groups. Although usually thought of fixed-time, interval control can also incorporate actuation. Interval timing is common in Europe.
- *Phase* timing, as the name suggests, focuses on the scheduling of phases in repeating structures called *rings*. A phase is a timing element associated with a green duration and a clearance time. A ring is an ordered list of phases that the controller serves in a cyclic fashion. Phases within the same ring are not typically compatible. However, other rings operating at the same time control phases that are compatible. In addition to phases, there exist another set of timing elements called *overlaps*, which are based on logical combinations of phase states. Pedestrian phases may be attached to vehicular phases or may exist on their own as exclusive pedestrian phases. Phase timing is common in North America.

The performance measure methodology of this monograph relies heavily on the concept of a *service instance*. A service instance is defined as a duration of time when a signal indication passes through a “Stop” state (red), a “Go” state (green), a “Clearance” state (yellow), and back to the “Stop” state. The duration of a service instance begins with one end of green and ends with the next end of green. This definition is based on the fact that vehicles arriving for service on a particular movement typically cannot advance past the intersection until

TABLE 3.1
Operating modes for signal systems broken down by local control and system control options.

Local Control	System Control	
	Fixed Cycle	Variable Cycle
Fixed Time	Fixed time coordinated	—
Actuation on Some Movements	Semi-actuated coordinated	Semi-actuated noncoordinated
Actuation on All Movements	Fully actuated coordinated	Fully actuated noncoordinated
Actuation with Prediction	Fixed cycle adaptive	Variable cycle adaptive

they have received a green indication. The duration of the green time, the number of vehicles counted during the service instance, and the percentage of time that the detectors in the associated lanes are occupied are three fundamental pieces of information from which performance measures can be calculated.

In a phase-based controller, each service instance can be defined by a chain of events, as illustrated in Figure 3.1. First, a *call* for service for the phase is registered in the controller. When each phase terminates, the controller makes a *next phase* decision that is influenced by the sequence in effect, phase compatibility rules, and other configuration settings. Over time, the controller serves all of its phase calls.

The duration of the green time provided to the phase depends on a variety of programmed settings. The green time must be long enough to serve the minimum green and usually any associated pedestrian walk and clearance times. The phase can be extended past the minimum time (by detection or recalls) up to a maximum. This maximum is determined either by the maximum timer (when not in coordination), or by the split timer (when in coordination).

- If all of the vehicle extension timers associated with the phase have expired, the phase will terminate in what is called a *gap-out*.
- If the maximum timer expires, terminating the phase, this is a *max-out*. The maximum timer begins counting down when the first call arrives for service of a different, conflicting, phase.
- If the split timer expires, terminating the phase, this is a *force-off*. The split timer is based on the background coordination cycle, ensuring that the coordinated phase becomes green on schedule.

Figure 3.2 shows some of the observable events associated with the service instance of a phase.

- The vehicle phase indications are the output of the signal controller. The most important events are the beginning of green (BOG) and the end of green (EOG). A phase is active whenever any of its intervals are timing—including the yellow and red clearance intervals. Otherwise, the phase is inactive. The green interval indicates when the movement is protected or when it has the right-of-way in the intersection. For example, for left-turn phases, “green” refers only to the time when the green arrow is displayed.

- Vehicle detectors provide the ability to measure demand for the phase. Figure 3.2 illustrates a pattern that is typical of a stop bar detection zone programmed to call and extend the phase, using equipment is able to measure both *presence* and *count* of vehicles crossing the zone. Typically, a stop bar detection zone will be as long as a few passenger cars, meaning that more than one count might be registered prior to green.
- Pedestrian phase indications and detector activity support similar measurements of the pedestrian demand. At the time of writing, pedestrian detectors are in development, but have seen little field deployment. Currently, the only measurement of pedestrian demand available at most intersections is the pedestrian pushbutton. In the United States, there is a walk indication, followed by a pedestrian clearance indication (“flashing don’t walk,” or FDW), and a steady “don’t walk” (DW) indication.
- Separate indications for bicycles are in use at some intersections. The most substantial differences in bicycle timing are the different minimum greens and clearance times.

Although these concepts are illustrated for a phase, they also extend to overlaps, or to signal groups in interval-based timing. Each signal-controlled movement has a sequence of intervals consisting of beginnings and endings of green, and detection within the movement lanes represents a measure of the demand for the movement. Although the internal timing rules may vary, the elements forming the basis of performance measures would be the same.

In the United States, a very common phase configuration is the *dual-ring, eight-phase* scheme (28), an example of which is illustrated in Figure 3.3. This diagram shows the progression of phases over time; when the end of the diagram is reached, the controller loops back to the beginning. The two rows represent the two rings. The controller can serve one phase per ring at any time. Each phase *group* is a collection of phases for which each phase is compatible with all of the phases in the same group in the other rings. For example, phase 5 is compatible with phases 1 and 2, but not with phase 6. Dual-ring timing is more flexible than single-ring timing, because it allows some concurrent phases to terminate at different times.

The two groups are separated by “barriers”; phase group 1 (containing phases 2 and 6) typically represent the “major” road, and group 2 typically represents the minor road. All phases currently in service must be in

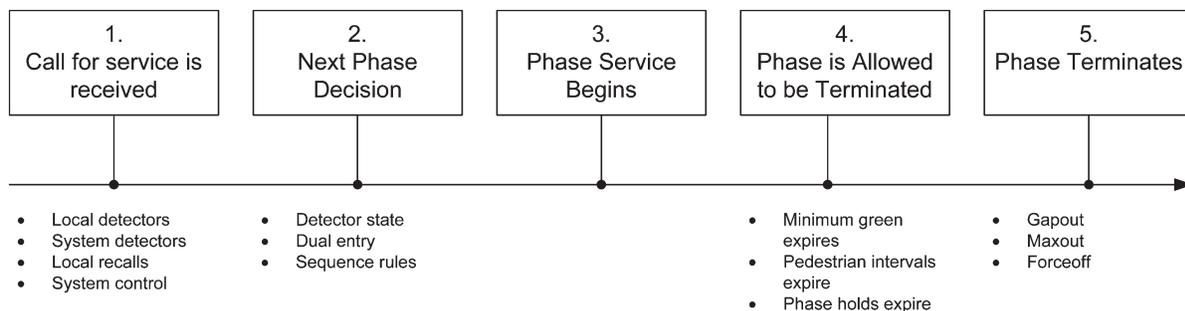


Figure 3.1 Events in local signal phase timing.

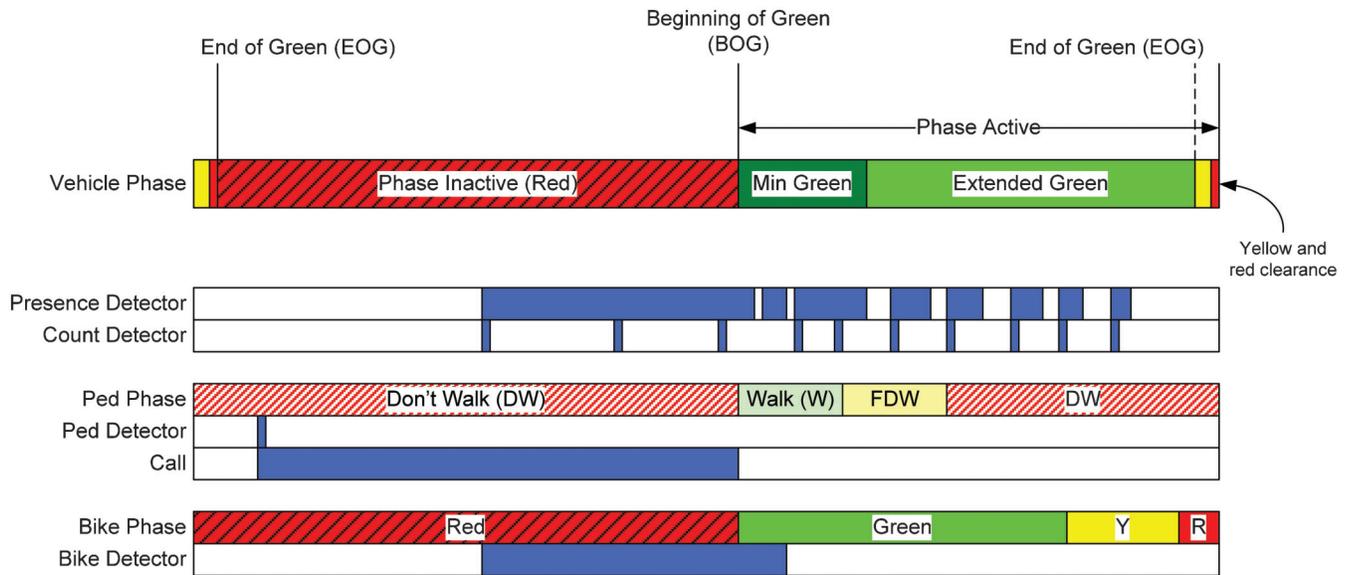


Figure 3.2 Basic observable events for monitoring phase performance.

the same phase group. When any ring reaches a barrier, it must wait for all other rings to reach the same point, after which they all may “cross the barrier” together. In many modern controllers, the barrier concept has been replaced with a *compatibility matrix*, which allows somewhat more flexibility. In any case, at most intersections, there is one group of vehicular movements coming from a “major” street that is altogether incompatible with the movements of a “minor” street.

The flexibility of the eight-phase scheme has contributed to its popularity in the United States. Even when only one ring is needed, such as at a two-phase signal, it is common for the phase numbering to nevertheless be based on the eight-phase template (e.g., using phase 2 for the major road and phase 4 for the minor road). Although it is conventional to number the major movements as 2 and 6, any phase number could be potentially assigned to any particular movement, and any phase may be coordinated. Additionally, the actual phase sequence can be completely different from that shown in Figure 3.3, but the numbering of the

movements often follows the eight-phase scheme as a template.

Apart from phases, the other type of signal output is that of an overlap. Overlaps are logical combinations of the states of “parent phases.” Whenever a parent phase is in service, the overlap will also be in service. One use of overlaps is to control right-turn arrows, as illustrated in Figure 3.4. Overlap A has the parent phases 2 and 3. When phase 2 is green, overlap A also turns green. The controller terminates phase 2, making a next-phase decision. Because phase 3 (another parent of overlap A) is selected, overlap A remains green as phase 2 goes into clearance. At the end of phase 3, the next phase is *not* one of its parent, so overlap A terminates with phase 3.

Similar transitions can be seen for the other right turns. It is also possible to add “negative” elements to an overlap, such as preventing the arrow from coming up during a pedestrian phase, as illustrated in Figure 3.5. In this example, the overlap green is inhibited while the pedestrian indications of phase 2 are active. The overlap green can also be made to lead

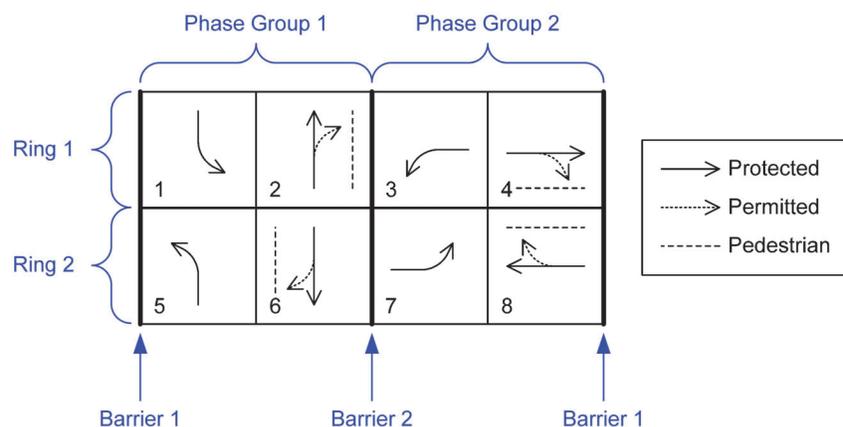


Figure 3.3 Standard configuration for dual-ring, eight-phase operation (north-south major road).

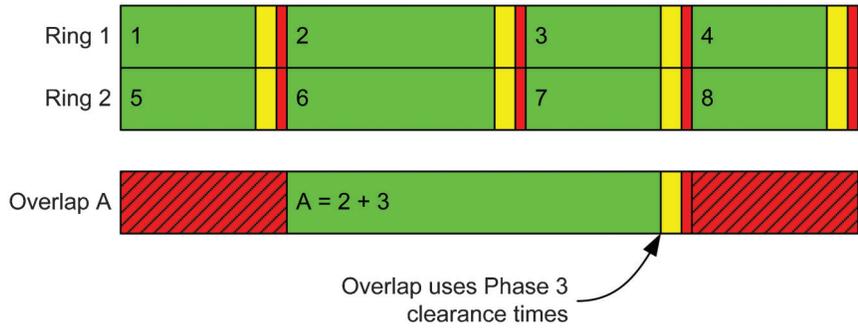


Figure 3.4 Example overlaps for a right-turn arrow indication.

or lag the greens of the parent phases, and pedestrian overlaps can be set up to carry over a walk indication from one phase to another (such as at intersections with two-stage pedestrian crossings). With a little creativity, overlaps can be used to create many different types of timing requirements for complex intersections or desired timing arrangements. It is possible to run an intersection entirely with overlaps, using different arrangements of parent phases to establish the desired operation. For example, pedestrian phases are never incompatible with another, and it is possible to improve their efficiency with resourceful uses of overlaps (29).

3.3 Linking Vehicle Detection and Signal Output

The demand for a movement, and for the phase or overlap that controls it, can be measured by counting the number of vehicles observed on lanes dedicated to that movement. A key assumption made in the performance measure methodology is that vehicle counts that accrue during red represent vehicles that are likely served in the subsequent green. That green interval could be protected or permitted, depending on the intersection and phase/overlap configuration. Four types of counting intervals are illustrated in Figure 3.6.

- Figure 3.6a illustrates the counting interval definitions for a protected-only movement. The counting interval is

defined by the two successive end of green (EOG) events; all of the vehicles counted during the red interval are expected to be served during the protected green.

- When a movement also has a permitted interval (Figure 3.6b), the counting interval is expanded to include the permitted interval, which is defined by the beginning of permitted green (BPG) and the end of permitted green (EPG). In this case, the protected and permitted intervals are separated by an intermittent red (IR) interval. Counts during IR are also included, because they are likely served during the subsequent permitted interval.
- Figure 3.6c shows the situation when the BPG is controlled by an adjacent through movement that receives a green before the protected left ends. In this case, the permitted green begins immediately upon the EOG, and there is no IR. This situation would be typical of flashing yellow arrow (FYA) operations, if the FYA interval commences immediately after the end of the protected interval.
- Finally, Figure 3.6d shows the situation where the turn is permitted-only. In this case, the counting interval is defined by the EPG only.

Figure 3.7 further illustrates some definitions affecting protected-permitted left turns. In Figure 3.7a and Figure 3.7b, the protected green g_{PROT} for phase 7 is identical and is based on the BOG and EOG of the protected phase.

- With a flashing yellow arrow (FYA), as in Figure 3.7a, the permitted interval is determined simply by the

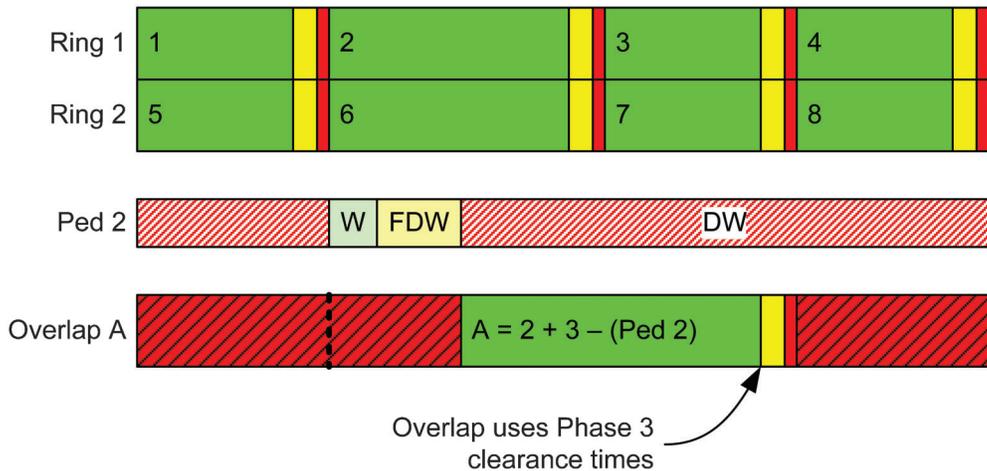


Figure 3.5 Example of a negative overlap.

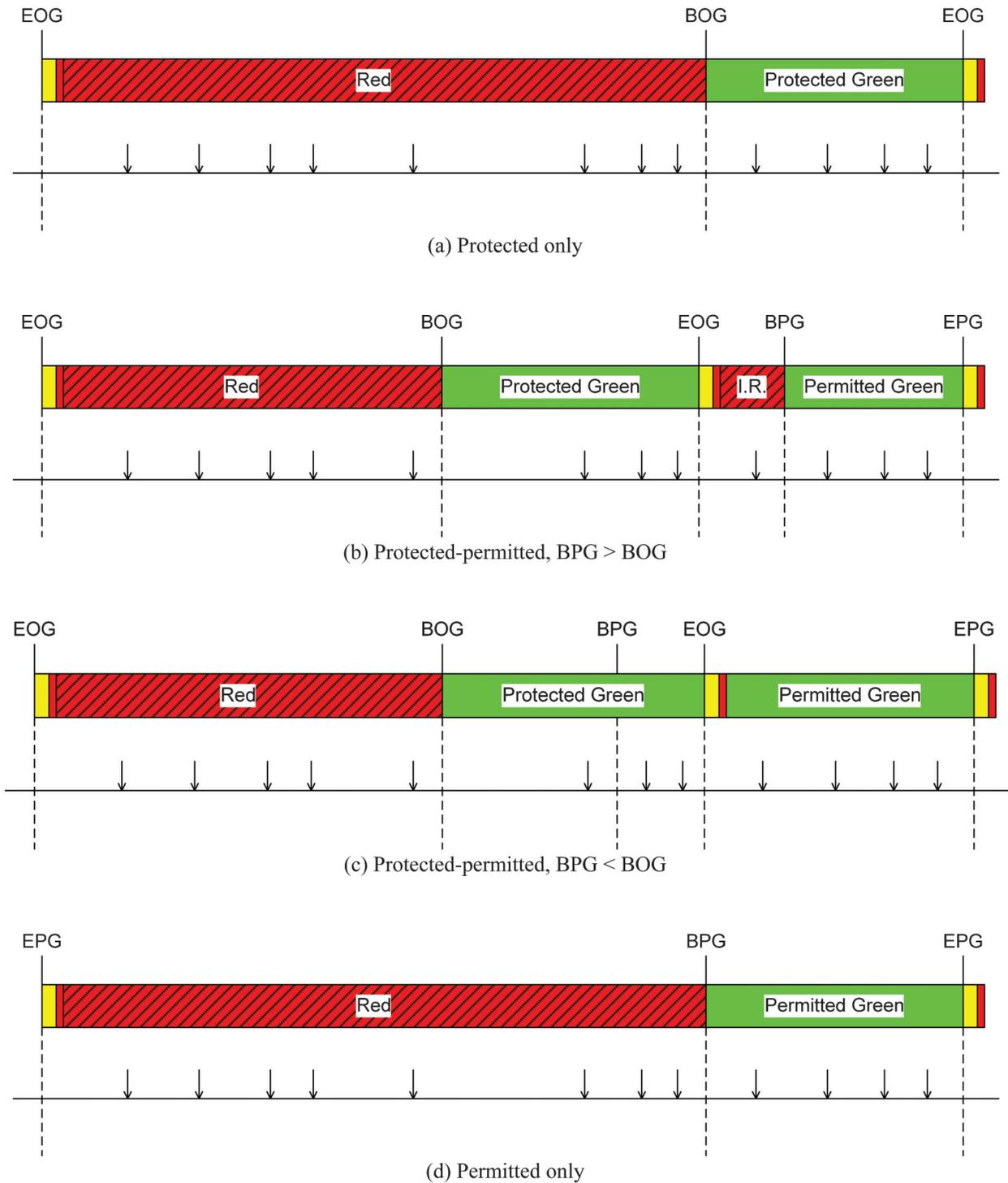


Figure 3.6 Counting intervals.

duration of the FYA indication, g_{FYA} . Unless a delay is programmed, this coincides with the duration of the opposing through movement. In any case, the beginning and end of the FYA will be directly measurable from events associated with the same phase or overlap.

- With a five-section head, as in Figure 3.7b, the permitted portion of the left turn is controlled by the *adjacent through green* (g_{AT}). As shown in Figure 3.6c, it is possible for the adjacent through to turn green before the

end of the protected left. In that case, the permitted portion is equivalent to the *opposing through* (g_{OT}). The duration of the permitted interval is the minimum of the two through movement greens:

$$g_{PERM} = \min(g_{AT}, g_{OT}). \quad (3.1)$$

This is explained more closely in Figure 3.8. If $g_{AT} > g_{OT}$, the permitted portion of phase 7 is equal to g_{AT}

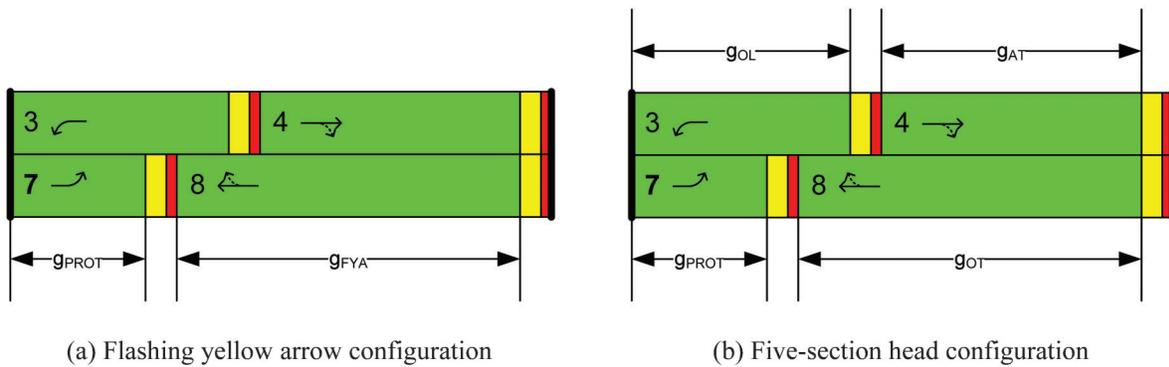


Figure 3.7 Permitted left-turn interval definitions.

(Figure 3.8a). If $g_{AT} < g_{OT}$, the permitted portion of phase 7 is equal to g_{OT} (Figure 3.8b).

3.4 Signal Cycles

This document defines a signal *cycle* as an interval of time in which the signal serves all phases in all phase groups (see Figure 3.3) for which there is demand. This may occur within a prescribed background cycle length, as during coordination, or the cycle length could vary, as during fully actuated operations. When phases 2 and 6 are the coordinated phases, the cycle length can be defined as the amount of time between subsequent crossings of barrier 2 (see Figure 3.3). This means that, within a signal cycle, the right-of-way has transitioned from the “major” road group to the “minor” road group, and back again. Each phase has had an opportunity for service. This definition does not vary with the lead/lag configurations of the coordinated phases and left-turn partner phases (30).

To compile performance measures, the service instances of phases and overlaps are mapped to cycles according to their BOG times. This is illustrated in Figure 3.9, which shows a phase timeline and the

mapping of service instances into the timeline. Because of phase actuation, it is not always true that the number of cycles within any given time period will equal the number of service instances of a phase or overlap. In fact, there might be multiple service instances belonging to a single cycle, if phase re-service is allowed.

In Figure 3.9, phase 1, a protected-only left turn, is omitted in cycles 101 and 102; the counting interval associated with its 84th service instance maps into cycle 103, since that is when its BOG occurs. Phase 2 is a protected through movement that occurs in every cycle. Its BOG lines up with each cycle in this example. Phase 7 is a protected-permitted left turn that has at least one protected or permitted interval in every cycle shown in the figure. The index time is the BOG when a protected interval exists. During cycle 102, the protected portion of phase 7 is omitted, meaning that it has only a permitted phase. The BPG then becomes the index time.

A phase (or overlap) **service instance** is defined as a *preceding effective red interval* and a *subsequent effective green interval*. The reason for defining a phase instance in this way is that all vehicles that arrive during the preceding red must exit during the following green interval.

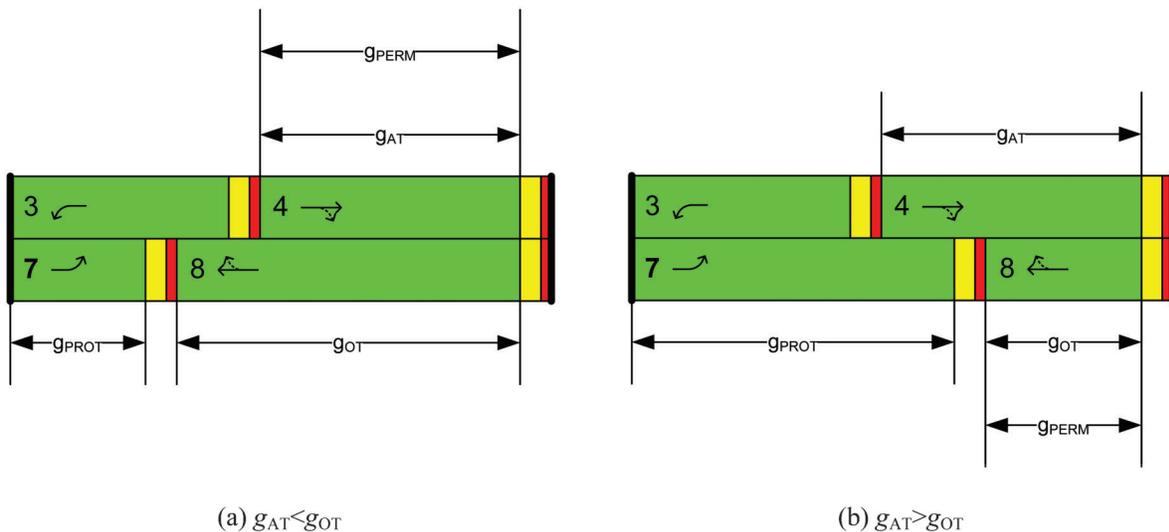


Figure 3.8 Permitted left-turn definitions for five-section head configuration.

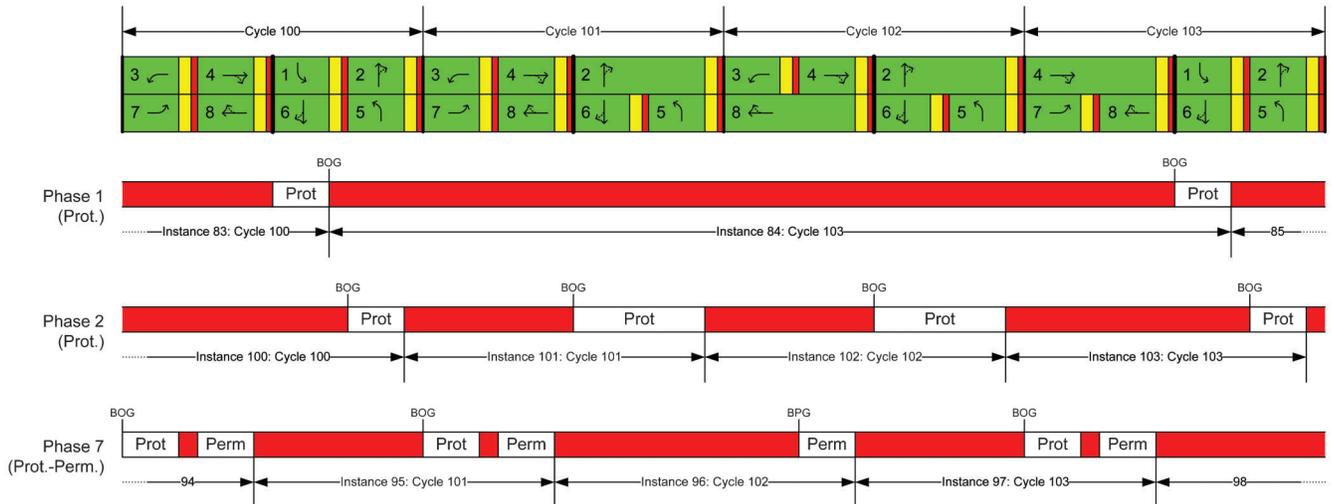


Figure 3.9 Relationship between service instances and cycles.

A **signal cycle** is defined as the time between two successive *barrier crossings* that occur at the end of the major coordinated phase group.

There is a *one-to-many* relationship between cycles and phase/overlap service instances: a given cycle might contain any number of instances (most often 0 or 1, but sometimes 2 or more).

3.5 Actuation

There are a number of ways in which local controllers determine the timing of phase green states from detection. This section describes some common methods of phase actuation.

Figure 3.10 illustrates how a gap timer works. When a phase becomes green, it first passes through its initial green, which includes the minimum green time as well as other optional added initial time. Depending on the controller settings, the gap timer may begin timing at the very beginning of green or at the end of the minimum and initial. The gap timer will not start unless

a call for a conflicting phase is registered. Once the gap timer has started, its value is held at the “vehicle extension” or “passage” time while the detector is occupied. The timer begins counting down to zero when the detector becomes unoccupied. When zero is reached, the phase *gaps out* and is allowed to be terminated. If the detector becomes occupied again, the gap timer can reset to the passage time if the “simultaneous gap” feature is active. “Gap reduction” timing causes the passage time to be reduced after an amount of green time has elapsed, called the “time before reduction.” The passage time is linearly reduced to the minimum gap over the “time to reduce.”

“Added initial” timing is compatible with advance or setback detection. During the red interval, detector actuations accrue an initial green time that can extend the minimum green. The idea is to measure the growing queue and ensure that enough green time is provided to serve it. Figure 3.11 illustrates how it works. First, a certain number of actuations is surpassed (“actuations before”) before added initial time begins accruing. Thereafter, each actuation adds an amount of time

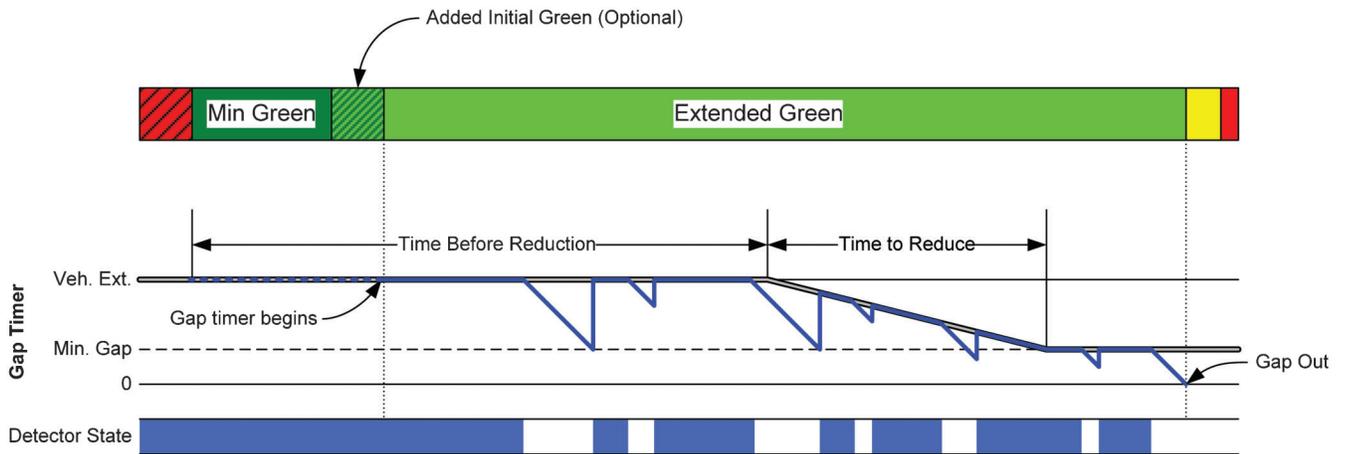


Figure 3.10 Variable gap timing.

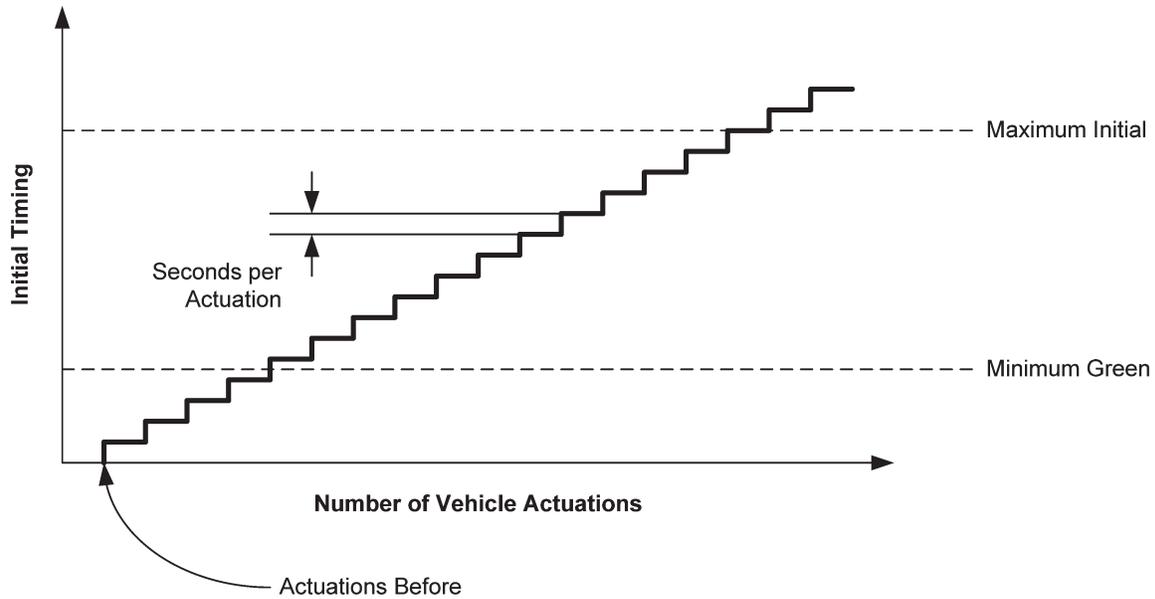


Figure 3.11 Added initial timing.

(“seconds per actuation”) to the initial green, which is never less than the minimum green, yet does not exceed the “maximum initial.” This type of timing logic is necessary when there is no stop bar detection zone (the alternative is an arbitrarily long minimum green).

Phases with sustained heavy demand (or a failed detector) will not gap-out, but instead will be terminated by the max timer (in fully actuated operations) or by the split timer (in coordinated operations when the max timer is inhibited, or when the max time is greater than the split). Figure 3.12 illustrates the behavior of the max timer. The max timer for a phase begins to count down after the first call for a conflicting phase is registered in the controller. If that call expires, the timer returns to the max value. If the call persists and the max timer reaches zero, then the phase will *max-out* and terminate.

Some controllers can adjust maximum times according to phase termination, which is based on the idea

that successive max-outs indicate a need for additional capacity. This is called “dynamic maximum” or “adaptive maximum” timing. Figure 3.13 illustrates dynamic maximum timing over several cycles; the phase maxes out during cycles 2–10 and gaps out in cycles 1 and 11–12. In the first three cycles, the max timer is set to its initial maximum time (Max. 1). After two max-outs (the minimum number, in this example), the max timer is increased by the “dynamic step.” This process continues until the eighth cycle, at which time the max timer has reached its ultimate maximum (“dynamic maximum”) and cannot be increased further. The max time reverts to its initial value after the phase gaps out in cycle 11.

The implementation of these phase timing parameters varies by controller, as illustrated by Table 3.2. This table provides the equivalent control settings for seven different signal controllers representing a variety

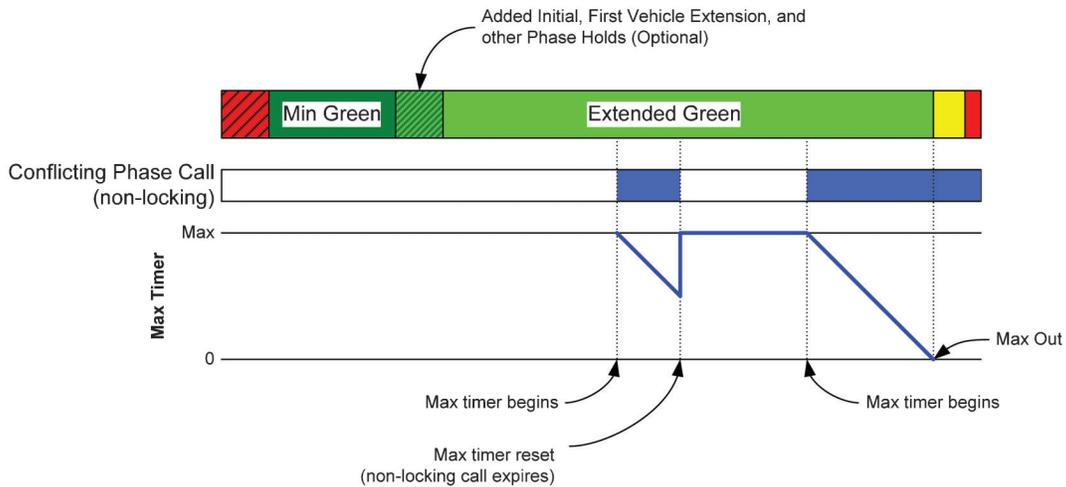


Figure 3.12 Phase maximum timing (fully actuated operations).

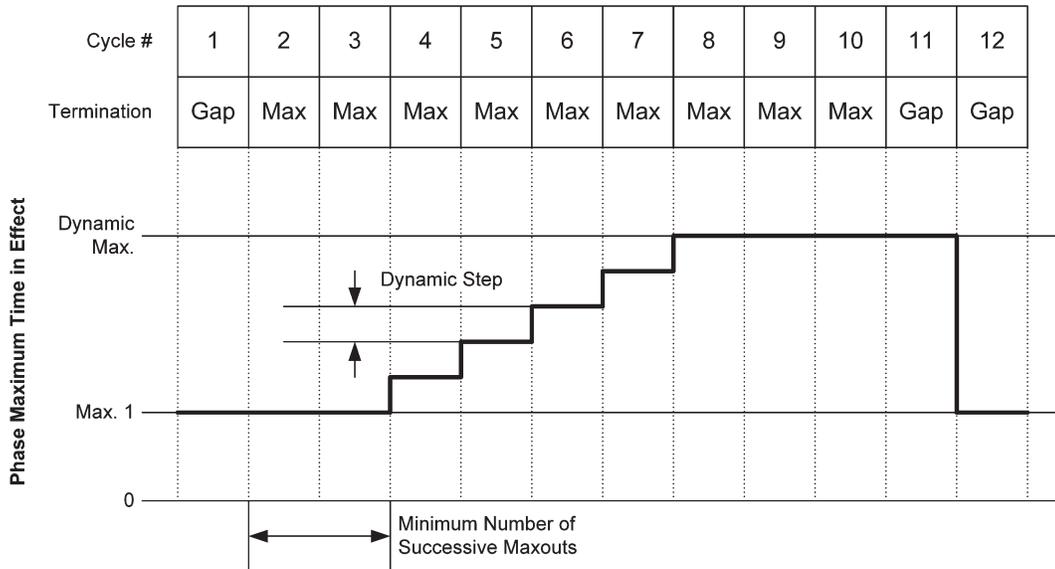


Figure 3.13 Dynamic maximum timing.

of vendors and controller ages. All of the controllers have common features such as minimum green time, yellow times, walk times, etc. There are some features, such as bike minimum green, that are common to a subset of controllers; others have unique features. This discussion will stop short of explaining each parameter. This example illustrates the complexity of the signal control problem once it is considered how the timing concepts are actually implemented (31).

3.6 Signal Coordination

An important objective of signal systems operation is to progress traffic along routes in the road network. This means scheduling the green times of adjacent intersections to allow arriving vehicles to pass through as many intersections as possible without stopping. Conventional coordination uses a common cycle length throughout the system to maintain such a schedule (28). Some adaptive systems predict or track platoons through the system or schedule green bands in advance (14). Some proposed algorithms may be able to accommodate platoons (32,33) or yield dynamic signal coordination as an emergent phenomenon (27,34). This discussion focuses on conventional coordination methods, since the central concepts apply to all methods of signal coordination.

Figure 3.14 shows a time-space diagram in which the y -axis represents distance along a coordinated street and the x -axis represents time. Each intersection shows a timeline of the signal state for a through movement on the coordinated street. A green band is shown as the shaded area that connects the green states of subsequent intersections; the slope of the boundaries represents the speed at which the first vehicle leaving Int. 1 is likely to travel to arrive on green at the subsequent intersections.

Vehicle speeds are largely determined by the road's geometric properties and the speed limit. Although it usually cannot be changed by the engineer, vehicle speed is important to coordination, since it controls the travel time to neighboring intersections. Longer distances between intersections and greater variability in speed lead to more variability in travel times. Variability in travel time causes platoons of vehicles to disperse over the link. The greater the distance between two signalized intersections, the less likely it is that coordination will benefit the arrivals on the two ends of the link. The presence of driveways, parking, and other roadside elements may also contribute to platoon dispersion.

In conventional coordination, a *split* is the proportion of the cycle time that is assigned to each particular phase by the local controller. Splits are expressed in percentage of cycle or number of seconds and are inclusive of the green time, yellow clearance, and red clearance time programmed for the phase. The minimum split must be sufficient to display these indications, as well the walk and pedestrian clearance intervals associated with the phase, and all of the clearance phases as well.

The *local zero* of each intersection is a reference point that repeats regularly in each cycle. The time between two subsequent local zeros is the cycle length. All of the intersections in the system have the same cycle length, with the possibility of some intersections using double or half-cycles. Coordination is established by creating a time difference, called the *offset*, between the local zero and a "system zero," a reference point for the entire system. The offset value programmed in the controller is always the offset with respect to the system zero. It is also possible to define a relative offset between any two intersections in the system. Figure 3.14 shows the

TABLE 3.2
Phase timing parameters in various controllers.

Phase Timing Aspect	Econolite ASC/3		Econolite ASC/2		Econolite Oasis 2070		Fourth Dimension D4		Intelight X-1		Naztec MC682		Peak ATC		Siemens SEPAC	
	MIN GRN	BK MGRN	MIN GRN	BIKE GRN	Minimum Green	Minimum Green 2	MinGrn	BikeMG	MinGrn	MIN GRN	MIN GRN	MIN GRN	MIN GRN	MIN GRN	MIN GRN	MIN GRN
Minimum Green	—	—	MIN GRN	—	Minimum Green	Minimum Green 2	—	—	—	—	—	—	—	—	—	—
Minimum Green 2	—	—	—	BIKE GRN	Alt. Minimum Green	—	—	—	—	—	—	—	—	—	—	—
Alternate Minimum Green	BK MGRN	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Delayed Green	DLY GRN	—	—	—	Walk Advance	—	—	—	DlyGrn	—	—	—	—	—	—	—
Early Walk	—	—	—	—	Walk Time	—	ElyWlk	—	—	—	—	—	—	—	—	—
Delayed Walk	—	—	—	—	Walk Delay Time	—	DlyWlk	—	DlyPed	—	—	—	—	—	—	—
Walk	WALK	WALK	WALK	WALK	Walk 1	Walk 2	Walk	Walk	Walk	WALK	WALK	PED WALK	—	—	—	WALK
Walk 2	WALK2	—	—	—	Walk 2	Alternate Walk	Walk2	Walk2	—	WALK 2	WALK 2	—	—	—	—	—
Alternate Walk	—	—	—	—	Alternate Walk	—	—	—	—	—	—	—	—	—	—	—
Walk Maximum	WLK MAX	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Pedestrian Clearance	PED CLR	—	PED CLR	—	Ped Clearance	—	PedClr	—	PedClr	—	—	PED CLR	—	—	—	PED CLR
Pedestrian Clearance 2	PD CLR2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Pedestrian Clearance Maximum	PC MAX	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Alternate Pedestrian Clearance	—	—	—	—	Alternate FDW	—	—	—	—	—	—	—	—	—	—	—
Pedestrian Carryover	PED CO	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Negative Pedestrian Hold-off	—	—	—	—	—	—	NegPed	—	—	—	—	—	—	—	—	—
Solid Don't Walk	—	—	—	—	—	—	SolidW	—	—	—	—	—	—	—	—	—
Audible Pedestrian Disconnect Time	—	—	—	—	—	—	APDisc	—	—	—	—	—	—	—	—	—
Gap/Passage/Vehicle Extension	VEH EXT	—	VEH EXT	—	Gap 1	—	VEExt	—	Passage	—	GAP, EXT	—	PASSAGE	—	PASS/10	—
Gap/Passage/Vehicle Extension 2	VH EXT2	—	VEH EXT2	—	Gap 2	—	—	—	—	—	—	—	—	—	—	—
Maximum Green 1	MAX 1	—	MAX1	—	Max 1	Max 1	Max 1	Max 1	Max 1	MAX 1	MAX 1	MAXIMUM 1	—	MAX # 1	—	—
Maximum Green 2	MAX 2	—	MAX2	—	Max 2	Max 2	Max 2	Max 2	Max 2	MAX 2	MAX 2	MAXIMUM 2	—	MAX # 2	—	—
Maximum Green 3	MAX 3	—	MAX3	—	Max 3	Max 3	Max 3	Max 3	Max 3	MAX 3	MAX 3	MAXIMUM 3	—	—	—	—
Det Fail Max Green Time	—	—	DET MAX	—	—	—	—	—	—	—	—	—	—	—	—	—
Dynamic Maximum	DYM MAX	—	—	—	Dynamic Max	—	MaxExt	—	DynMax	—	—	—	DYNAMIC MAX	—	—	—
Dynamic Step	DYM STP	—	—	—	Dynamic Max Adj	—	—	—	MaxStep	—	—	—	DYNAMIC MAX	—	—	—
Yellow Clearance	YELLOW	—	YELLOW	—	Yellow Clearance	—	Yel	—	YelChg	—	YELLOW	—	YELLOW	—	YEL/10	—
Red Clearance	RED CLR	—	RED CLR	—	Red Clearance	—	RedClr	—	RedClr	—	RED	—	RED CLEARANCE	—	RED/10	—
Red Maximum	RED MAX	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Maximum Waiting Time	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Red Revert	RED RVT	—	RED RVT	—	*	—	RedRvt	—	RedRvt	—	*	—	RED REVERT	—	*	—
Actuations Before	ACT B4	—	ACT B4	—	—	—	—	—	—	—	—	—	—	—	—	—
Seconds per Actuation	SEC/ACT	—	SEC/ACT	—	—	—	Added	—	—	—	—	—	—	—	—	—
Maximum Added Initial	MAX INT	—	MAX INT	—	Max. Variable Initial	—	MaxInt	—	—	—	—	—	MAX INITIAL	—	MAX INI	—
Time Before Reduction	TIME B4	—	TIME B4	—	Time Before Red.	—	RedAft	—	—	—	—	—	SEC/ACTUATION	—	AINI/10	—
Cars Waiting Before Reduction	CARS WT	—	CARS WT	—	—	—	—	—	—	—	—	—	—	—	—	—
Step to Reduce	STPTDUC	—	TTREDUC	—	—	—	—	—	—	—	—	—	—	—	—	—
Time to Reduce	TTREDUC	—	TTREDUC	—	Time to Reduce	—	TTRed	—	—	—	—	—	—	—	—	—

TABLE 3.2
(Continued)

Phase Timing Aspect	Economite ASC/3		Economite ASC/2		Economite Oasis 2070		Fourth Dimension D4		Intelight X-1	Naztec MC682	Peek ATC	Siemens SEPAC					
	MIN GAP	MIN GAP	MIN GAP	CS MGRN	Minimum Gap	Reservice Green	MinGap	CSMin	CSMax	PmtGrn	PmtWlk	PmtPC	RtnGrn	MIN GAP	MINIMUM GAP	MGAP/10	
Minimum Gap	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Conditional Service Minimum Green	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Conditional Service Maximum Green	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Preempt Green	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Preempt Walk	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Preempt Pedestrian Clearance	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Preempt Return Green	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

*Not specified on a phase-specific basis.

actual offset of Int. 2 and its relative offset between Int. 1.

During coordination, the timing of individual phases is scheduled to occur at particular times in the cycle (with some flexibility), with the objective of eventually returning the right-of-way to the coordinated phases at the scheduled time in the cycle. One way of thinking about coordination is to consider that the signal always serves the coordinated phases by default; the noncoordinated phases are allowed to begin service during “permissive periods” (or “windows”). The permissive periods are structured such that no phase can start unless it can be terminated in time to keep the other phases on schedule. The individual phase durations are controlled by the split timer.

Figure 3.15 illustrates rules governing the termination of noncoordinated phases in conventional coordination. This illustrates the function of the split timer as well, which has two different types of behaviors depending on the force-off type.

- With *fixed* force-off, the split timer begins when the controller enters the portion of the background cycle assigned by the split parameters. In Figure 3.15a, phase 4 is able to start early owing to phase 3 being skipped. However, the split timer is frozen at the split value until the controller reaches the point in the cycle as determined by the programmed split times. Note that this scheme allows phase 4 to obtain a longer green time than its programmed split.
- With *floating* force-off, the split timer begins counting immediately upon the beginning of green. This has the effect of placing a “float” on top of the cycle for serving the phase that is guaranteed to close after the split has been served. Phase 4 cannot be served for any more time than its programmed split during an individual cycle. By doing that, floating force-off causes any cycle time that is given up by noncoordinated phases to be absorbed by the coordinated phase, which has different termination rules.

The termination of a coordinated phase is slightly more complicated. There must be a call for a conflicting phase, and that call must occur within its permissive period. The coordinated phase must also be able to terminate. This requires that the coordinated phase be past its yield point. Past the yield point, the coordinated phase behaves as an actuated phase, and it can either be terminated by a lack of detector activity (gap-out) or extend until its split timer reaches zero (force-off).

Figure 3.16 illustrates two cycles of operation for a coordinated phase. In this simplified example, there is assumed to be only one conflicting, noncoordinated phase. It is also presumed that there is a constant call for the coordinated phase. In the first cycle, the split timer begins counting down once the local clock enters the coordinated phase split. However, the phase does not terminate when the split timer reaches zero, because there is not a call for a conflicting phase at that time. Although a call appears later, the controller does not serve it, because the call is placed outside of the permissive period. Because there are no calls on any conflicting phases, the coordinated phase split timer

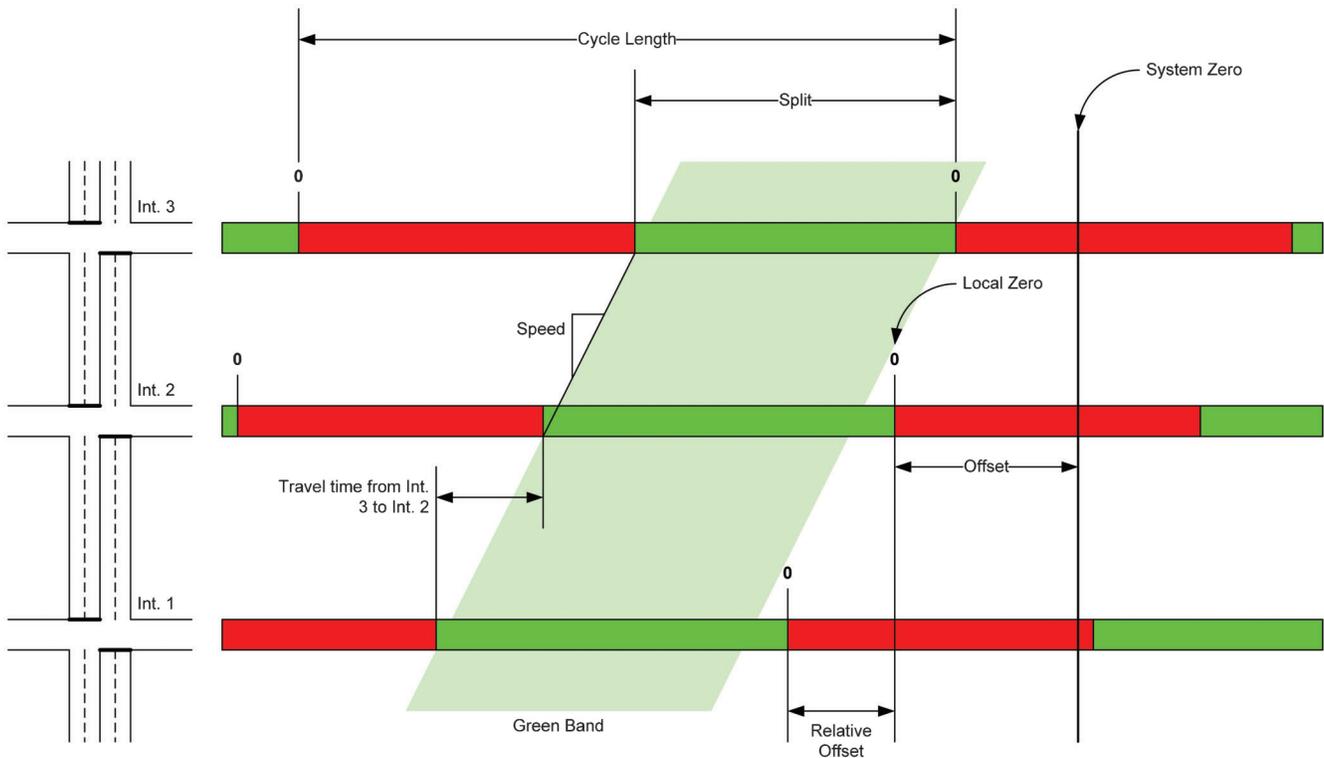


Figure 3.14 Signal coordination concepts.

resets to its maximum value, and the controller rests in green on the phase. Once the background clock enters the coordinated phase split, the timer begins counting. At the yield point, the permissive window opens for the next conflicting phase. It is now possible for the coordinated phase to terminate. However, in this example, the coordinated phase is extended. At the end of the split, however, the coordinated phase is forced off in order to serve the conflicting phase.

The timing illustrated in Figure 3.16 shows an event called the *yield point*. This defines the earliest time in a cycle when the coordinated phase may terminate. In some configurations, the coordinated phase is allowed to retain its entire split during each cycle. Using the “early yield” or “actuated coordinated” feature allows the coordinated phase to gap-out earlier, when detection is present on the approach. This type of operation is often favored by INDOT because of improvements in the dynamic reallocation of green time (35,36).

Figure 3.17 shows the difference between non-early yield coordinated phases (Figure 3.17a) and early yield coordinated phases (Figure 3.17b), and the resulting operation when coordinated phases 2 and 6 both have the ability to gap-out. Under conventional operation, Figure 3.17a, the coordinated phases cannot terminate early, and thus they continue timing to the end of the cycle. None of the other phases have access to the time that otherwise could have been yielded by phases 2 and 6. In contrast, with early yield coordinated phases (Figure 3.17b), the coordinated phases can end early, and other phases have an opportunity to start earlier

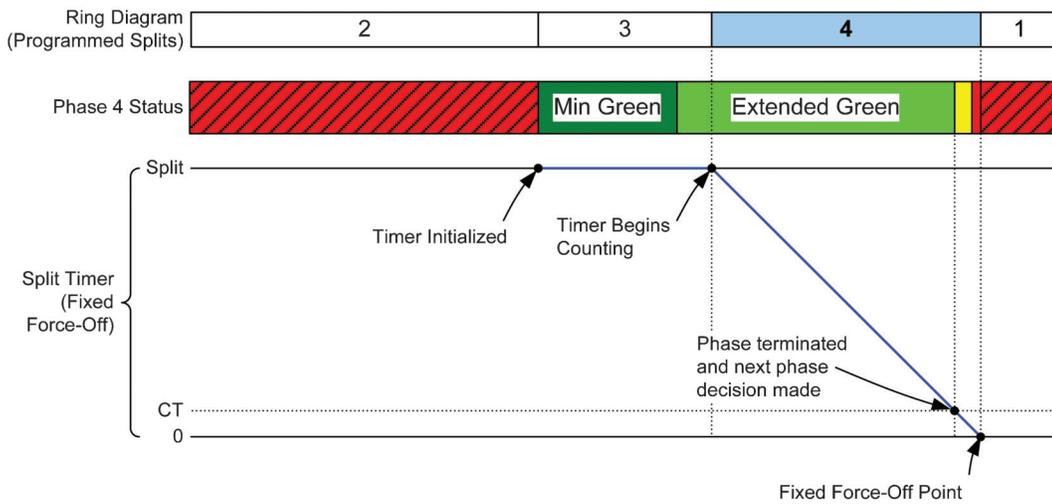
and to obtain more capacity. At the same time, the coordinated phase retains the possibility to extend its green and access the time if needed. In Figure 3.17b, the first cycle terminates early, while the second cycle retains the entire coordinated phase split.

An additional detail in signal coordination relevant to the programming of timing plans is the definition of the reference point, where the local zero exists in the cycle. Figure 3.18 illustrates a variety of possible reference points for a signal where coordinated phases 2 and 6 are set up in a lead-lag configuration. The implementation of reference points in various signal controllers is explained in Table 3.3.

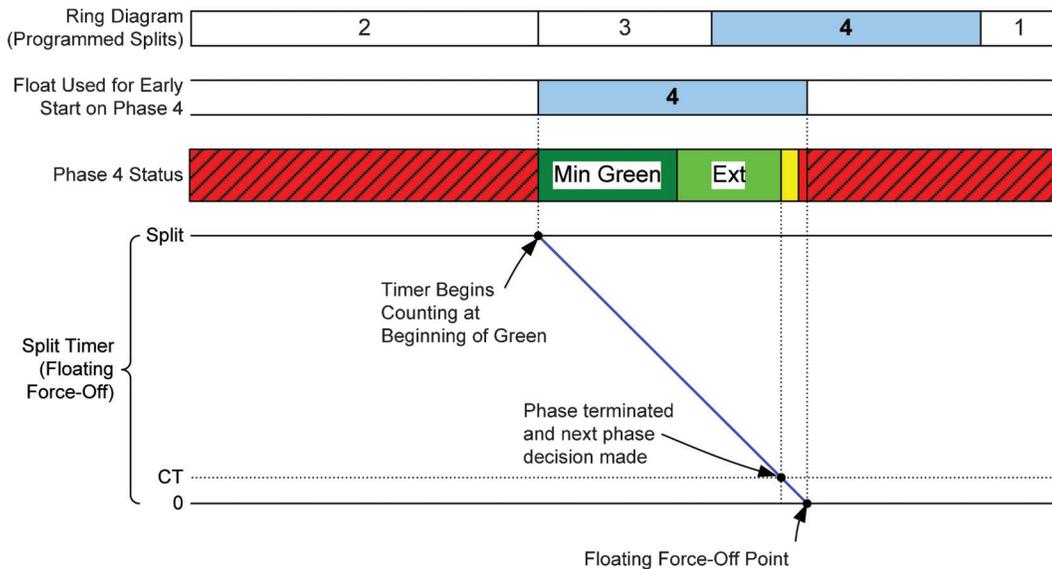
3.7 Preemption and Priority

Most signal controllers feature two special functions that enable right-of-way to be changed to accommodate special operating conditions, often serving a special mode other than vehicle, pedestrian, and bicycle movements. These are *preemption*, which causes the controller to jump to a desired state as soon as possible, and *priority*, which causes the controller to transition to a desired state faster than normal, but usually more gradually than preemption. Typically, preemption is used for emergency vehicles and adjacent railroad crossings, whereas priority is used for transit vehicles.

Figure 3.19 shows a typical timeline of a signal preemption event. First, the preempt input is activated by an external source. Optionally, a delay time can have the controller wait before entering the regular preemption



(a) Fixed force-off



(b) Floating force-off

Figure 3.15 Phase split timing (noncoordinated phase).

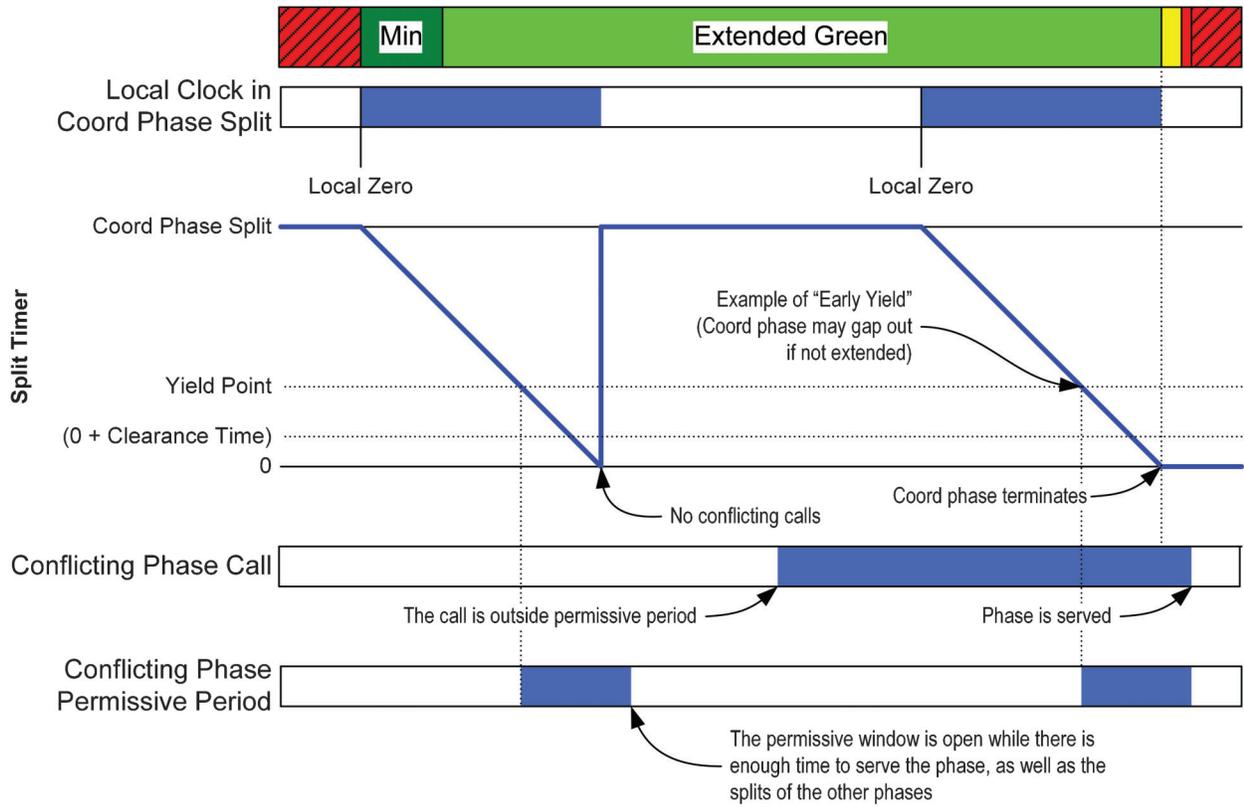
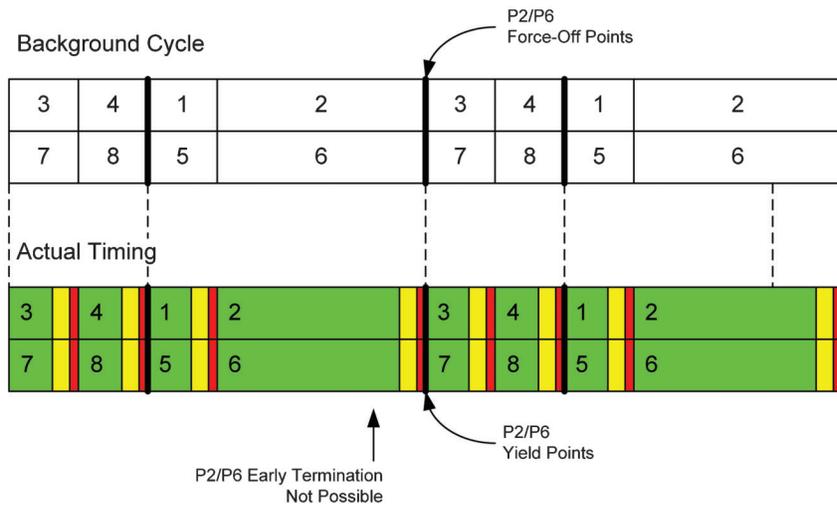
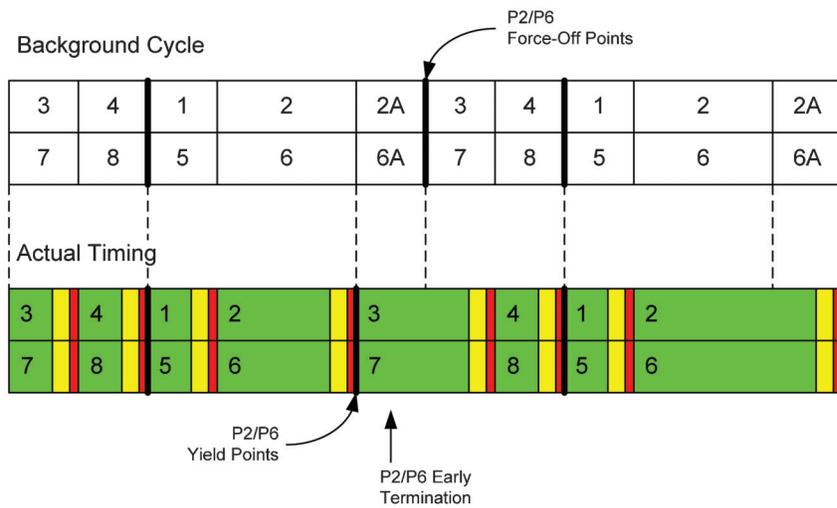


Figure 3.16 Phase split timing (coordinated phase).



(a) Coordinated phases that do not have an actuated portion



(b) Coordinated phases that have an actuated portion

Figure 3.17 Actuated-coordinated phases.

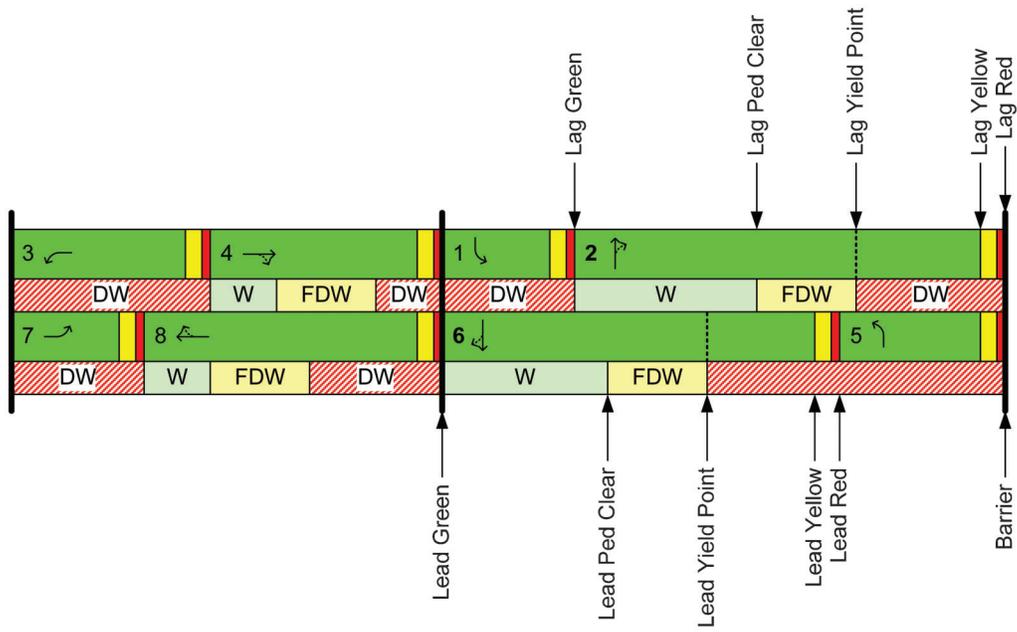


Figure 3.18 Coordination reference points.

sequence. The number of states within the preemption sequence varies by controller. Some controllers feature an entry state and a track clearance state, with the purpose of the latter being to clear vehicles from railroad tracks. A “limited service” state is common to all controllers, which is the desired signal state for the duration of the preemption. This might place the intersection in flash, dwell in a certain state, or allow the controller to continue cycling, with some phases omitted. On the termination of the preempt input, the controller transitions to an exit state (if exit phases are defined), and returns to normal operations.

Railroad preemption with a track clearance phase is slightly more complicated, because of the potential need to stage the track clearance interval with the railroad crossing gates. Figure 3.20 shows a timeline of these events from the perspective of the preemptor, the railroad

crossing lights and gates, and the track clearance phase. The controller preempt state becomes active at the onset of the advance preempt. After the advance preempt time (APT), the active warning time (AWT) begins with the onset of the crossing lights. After this point, the gates will descend, closing the path across the tracks. Before this time, the controller should enter the track clear green (TCG) state. The amount of time it takes to transition to this state is the right-of-way transfer time (RTT). Ideally, the TCG is extended until the crossing gates are fully down, although it is acceptable for it to be extended somewhat later. After this, limited service should begin.

Transit signal priority (TSP) works in a similar fashion. Figure 3.21 illustrates TSP operation. Transit vehicles (buses, trams, etc.) at an intersection activate a “check-in” input that alerts the controller to their arrival. A delay is optional, especially if check-in occurs

TABLE 3.3
Coordination reference points for selected controllers.

Reference Point	Controller			
	Econolite ASC/3	Fourth Dimension D4 Controller	Siemens/Eagle SEPAC PIM-177	Synchro/UTDF
Lead Green	Lead	LeadGrn	BEG	“TS2 First Green”
Lag Green	Lag	LagGrn	—	Green—“TS1 Style”
Lead Yield Point	Yield	LeadFO	—	—
Lag Yield Point	—	LagFO	—	—
Lead Yellow	Yellow	—	END	Yellow
Lag Yellow	—	—	—	—
Lead Red	—	—	—	Red
Lag Red	—	LagEnd	—	—
Lead Ped Clear/FDW	—	—	—	“170 Style”
Lag Ped Clear/FDW	—	—	—	—
Barrier	—	CordEnd	—	—

NOTE: The Peek ATC allows the start of any specific phase number to serve as reference point, although Lead Green is the default.

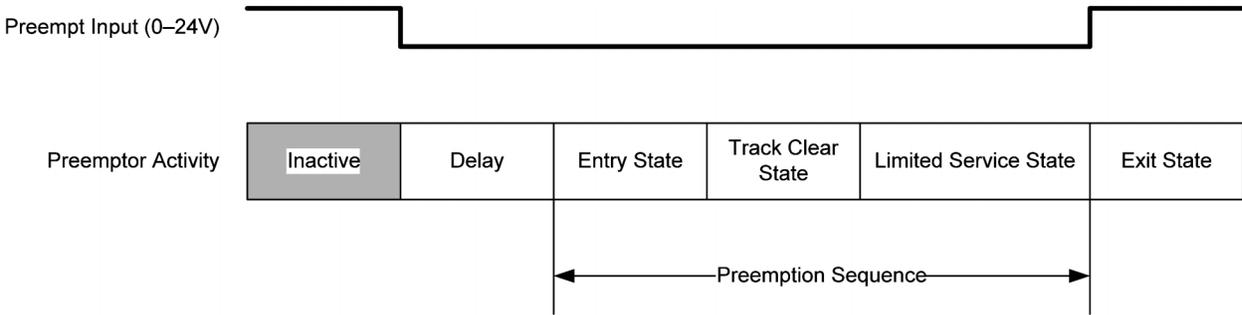


Figure 3.19 Common preemption events.

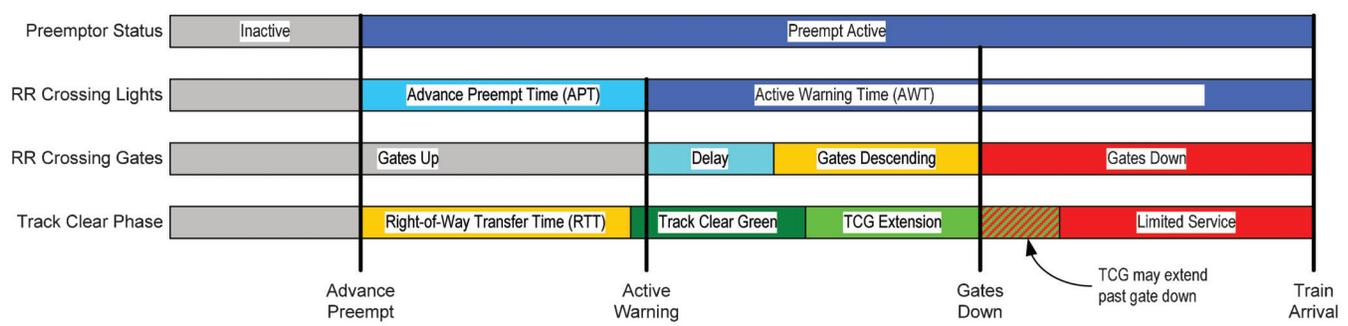


Figure 3.20 Events specific to railroad preemption.

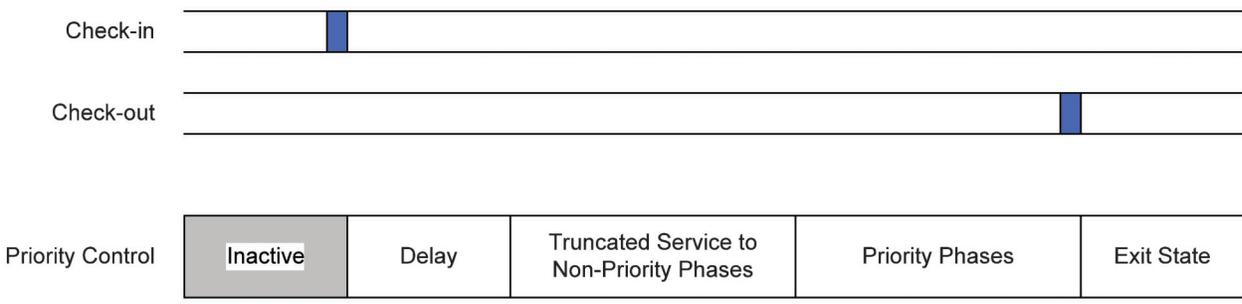


Figure 3.21 Transit signal priority.

far upstream. Unlike preemption, TSP seeks to more gradually reach the desired state, usually by providing some service to the nonpriority phases in sequence but reducing the amount of time given to them. Finally, the signal enters the priority phases, which are configured for the particular transit route for the given input. At some point, the transit vehicle passes a “check-out” point that ends the priority request (another option is for the request to expire after a specified amount of time), after which the controller may transition to a particular state or execute some other activity before returning to normal operations.

4. DATA METHODOLOGY

The data processing workflow is the framework around which the performance measures in the subsequent chapters are referenced. This chapter explains the workflow and the structure of the high-resolution

data. The discussion concerns the various time intervals relevant to signal operation and how these are grouped together to establish cycle-by-cycle performance measures. Finally, the chapter introduces the study intersection from which most of the subsequent performance measure example views are generated.

4.1 Performance Measure Methodology Overview

Figure 4.1 explains the workflow for developing operational performance measures from a *post-processing* perspective in which the data are analyzed after the analysis period. The workflow is closely tied in with the concept of service instances and cycles (see Figure 3.9). The first step, of course, is to obtain the data from the field. The two next steps are to extract the cycle times and phase intervals from the data, which provides the set of relevant time intervals that will anchor the performance measures. In most cases, the analysis

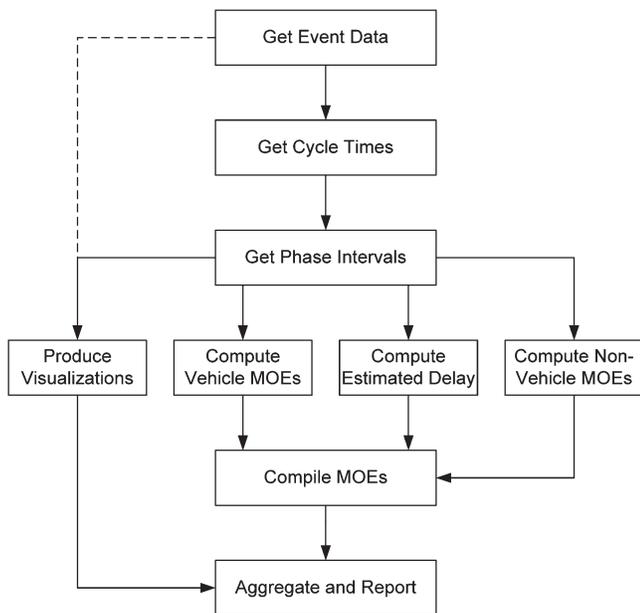


Figure 4.1 Performance measure analysis workflow.

period will be divided into a number of cycles, and there will be one service instance per cycle. When multiple service instances occur within a cycle, the properties of the individual service instances are combined to yield a performance measure for the entire cycle. This is done to enable aggregation across phases to calculate intersection-wide performance measures.

Once the cycles and phase intervals have been defined, it is possible to integrate the vehicle count event data to yield performance measures. Three separate tracks for performance measure development are identified in this methodology:

- *Produce visualizations.* The raw event data yields a number of graphical tools for characterizing aspects of signal performance. Examples include flow profiles and coordination diagrams, which are discussed further in Chapter 6.
- *Compute vehicle MOEs.* Vehicle counts may be compiled on any phase with a working count detector. These support cycle-by-cycle performance measures for a given lane, lane group, or phase. Operational performance measures for capacity utilization and the quality of progression are discussed in Chapters 5 and 6 respectively.
- *Compute estimated delay.* When upstream or setback/advance detectors are available, a record of vehicle arrivals at the intersection can be measured. This provides a means of estimating delay and queue length. This is discussed further in Chapter 6.
- *Compute nonvehicle MOEs.* Many intersections serve nonvehicle modes such as pedestrians or transit vehicles. The quality of detection varies considerably by mode and by location—from GPS trajectories of transit vehicles to pedestrian pushbutton actuation times. This monograph presents a variety of performance measures for non-vehicle modes in Chapter 7.

The next steps is to compile the cycle-by-cycle performance measures, yielding a series of data tables

that support reports based on aggregated data, enabling the analyst to “drill down” to lower-level data and examine areas of interest.

This monograph discusses two additional categories of performance measures for additional purposes beyond creating reports on signal operations. Chapter 8 discusses the development of performance measures for *equipment maintenance*, which is extremely important to ensure that operations can take place, yet is more oriented toward technicians. Chapter 9 discusses travel time data sets for *outcome assessment*, for before-after studies as well as in network-level assessments of transportation system travel time reliability.

4.2 Data Specification

In the past several years, the research team has worked in cooperation with INDOT and several signal controller manufacturers to establish a means of recording event data in a traffic signal controller. These data have taken on the name “high-resolution” because the events are written down in the highest time resolution of the controller (0.1 seconds), as opposed to volume-occupancy data, which are aggregated by minute. The first integrated data collector was described by Smaglik *et al.* (22). This was subsequently improved in several revisions. In 2011, the first stable version of the data specification for writing signal events was created by dialog between the research participants, and this has recently been published (23).

The events generated by the signal controller are outputted in sets of four bytes per event: one byte for the event code type, one byte for the event parameter (for signifying detector numbers and phases), and two bytes for the timestamp of when the event occurred. The event code is important for determining the type of activity the controller reports at a specific time—this could be phase initiation or termination, detection on/off, errors, etc. There are 256 possible activities reportable by the event code byte. Tables 4.1 through 4.9 detail the different types of events.

- *0–20. Active Phase Events (Table 4.1):* indicate any phase related status changes such as activation or termination.
- *21–30. Active Pedestrian Phase Events (Table 4.2):* indicate pedestrian-related phase status changes.
- *31–40. Barrier/Ring Events (Table 4.3):* indicate barrier and yellow permissive events.
- *41–60. Phase Control Events (Table 4.4):* indicate phase hold, call, and omit status changes.
- *61–80. Phase Overlap Events (Table 4.5):* indicate overlap status changes.
- *81–100. Detector Events (Table 4.6):* indicate detector activity and error status changes.
- *101–130. Preemption Events (Table 4.7):* indicate preemption status changes.
- *131–170. Coordination Events (Table 4.8):* indicate coordinated timing status changes, such as cycle length and split times.

TABLE 4.1
Active phase events (23).

Event Code ID	Name	Parameter	Description
0	Phase On	Phase # (1–16)	Set when NEMA Phase On becomes active, either upon start of green or walk interval, whichever occurs first.
1	Phase Begin Green	Phase # (1–16)	Set when either solid or flashing green indication has begun. Do not set repeatedly during flashing operation.
2	Phase Check	Phase # (1–16)	Set when a conflicting call is registered against the active phase. (Marks beginning of MAX timing.)
3	Phase Min Complete	Phase # (1–16)	Set when phase min timer expires.
4	Phase Gap-out	Phase # (1–16)	Set when phase gaps out, but may not necessarily occur upon phase termination. Event may be set multiple times within a single green under simultaneous gap-out.
5	Phase Max-out	Phase # (1–16)	Set when phase MAX timer expires, but may not necessarily occur upon phase termination due to last car passage or other features.
6	Phase Force-off	Phase # (1–16)	Set when phase force-off is applied to the active green phase.
7	Phase Green Termination	Phase # (1–16)	Set when phase green indications are terminated into either yellow clearance or permissive (FYA) movement.
8	Phase Begin Yellow Clearance	Phase # (1–16)	Set when phase yellow indication becomes active and clearance timer begins.
9	Phase End Yellow Clearance	Phase # (1–16)	Set when phase yellow indication become inactive.
10	Phase Begin Red Clearance	Phase # (1–16)	Set only if phase red clearance is served. Set when red clearance timing begins.
11	Phase End Red Clearance	Phase # (1–16)	Set only if phase red clearance is served. Set when red clearance timing concludes. This may not necessarily coincide with completion of the phase, especially during clearance of trailing overlaps, red revert timing, red rest, or delay for other ring terminations.
12	Phase Inactive	Phase # (1–16)	Set when the phase is no longer active within the ring, including completion of any trailing overlaps or end of barrier delays for adjacent ring termination.
13–20	Phase events reserved for future use	Phase # (1–16)	Set when NEMA Phase On becomes active, either upon start of green or walk interval, whichever occurs first.

TABLE 4.2
Active pedestrian phase events (23).

Event Code ID	Name	Parameter	Description
21	Pedestrian Begin Walk	Phase # (1–16)	Set when walk indication becomes active.
22	Pedestrian Begin Clearance	Phase # (1–16)	Set when flashing don't walk indication becomes active.
23	Pedestrian Begin Solid Don't Walk	Phase # (1–16)	Set when don't walk indication becomes solid (non-flashing) from either termination of ped clearance, or head illumination after a ped dark interval.
24	Pedestrian Dark	Phase # (1–16)	Set when the pedestrian outputs are set off.
25–30	Pedestrian events reserved for future use		

TABLE 4.3
Barrier and ring events (23).

Event Code ID	Name	Parameter	Description
31	Barrier Termination	Barrier # (1–8)	Set when all active phases become inactive in the ring and cross barrier phases are next to be served.
32	FYA – Begin Permissive	FYA # (1–4)	Set when flashing yellow arrow becomes active.
33	FYA – End Permissive	FYA # (1–4)	Set when flashing yellow arrow becomes inactive through either clearance of the permissive movement or transition into a protected movement.
34–40	Barrier events reserved for future use		

TABLE 4.4
Phase control events (23).

Event Code ID	Name	Parameter	Description
41	Phase Hold Active	Phase # (1–16)	Set when phase hold is applied by the coordinator, preemptor, or external logic. Phase does not necessarily need to be actively timing for this event to occur.
42	Phase Hold Released	Phase # (1–16)	Set when phase hold is released by the coordinator, preemptor, or external logic. Phase does not necessarily need to be actively timing for this event to occur.
43	Phase Call Registered	Phase # (1–16)	Call to service on a phase is registered by vehicular demand. This event will not be set if a recall exists on the phase.
44	Phase Call Dropped	Phase # (1–16)	Call to service on a phase is cleared by either service of the phase or removal of call.
45	Pedestrian Call Registered	Phase # (1–16)	Call to service on a phase is registered by pedestrian demand. This event will not be set if a recall exists on the phase.
46	Phase Omit On	Phase # (1–16)	Set when phase omit is applied by the coordinator, preemptor, or other dynamic sources. Phase does not necessarily need to be actively timing for this event to occur. This event is not set when phase is removed from the active sequence or other configuration-level change has occurred.
47	Phase Omit Off	Phase # (1–16)	Set when phase omit is released by the coordinator, preemptor, or other dynamic sources. Phase does not necessarily need to be actively timing for this event to occur. This event is not set when phase is added from the active sequence or other configuration-level change has occurred.
48	Pedestrian Omit On	Phase # (1–16)	Set when ped omit is applied by the coordinator, preemptor, or other dynamic sources. Phase does not necessarily need to be actively timing for this event to occur. This event is not set when phase is removed from the active sequence or other configuration-level change has occurred.
49	Pedestrian Omit Off	Phase # (1–16)	Set when ped omit is released by the coordinator, preemptor, or other dynamic sources. Phase does not necessarily need to be actively timing for this event to occur. This event is not set when phase is added from the active sequence or other configuration-level change has occurred.
50–60	Phase Control Events reserved for future use		

TABLE 4.5
Overlap events (23).

Event Code ID	Name	Parameter	Description
61	Overlap Begin Green	Overlap #*	Set when overlap becomes green. Do not set repeatedly when overlap is flashing green. Note that overlap colors are consistent to the GYR intervals resultant from the controller programming and may not be indicative of actual signal head colors.
62	Overlap Begin Trailing Green (Extension)	Overlap #	Set when overlap is green and extension timers begin timing.
63	Overlap Begin Yellow	Overlap #	Set when overlap is in a yellow clearance state. Note that overlaps which drive yellow field indications during a dwell state may be reported as green or inactive. (Common to mid-block signals)
64	Overlap Begin Red Clearance	Overlap #	Set when overlap begins timing red clearance intervals.
65	Overlap Off (Inactive with red indication)	Overlap #	Set when overlap has completed all timing, allowing any conflicting phase next to begin service.
66	Overlap Dark	Overlap #	Set when overlap head is set dark (no active outputs). The end of this interval shall be recorded by either an overlap off state or other active overlap state.
67	Pedestrian Overlap Begin Walk	Overlap #	Set when walk indication becomes active.
68	Pedestrian Overlap Begin Clearance	Overlap #	Set when flashing don't walk indication becomes active.
69	Pedestrian Overlap Begin Solid Don't Walk	Overlap #	Set when don't walk indication becomes solid (non-flashing) from either termination of ped clearance, or head illumination after a ped dark interval.
70	Pedestrian Overlap Dark	Overlap #	Set when the pedestrian outputs are set off.
71–80	Overlap events reserved for future use	Overlap #	

*Overlaps numbered as A = 1, B = 2, etc.

TABLE 4.6
Detector events (23).

Event Code ID	Name	Parameter	Description
81	Detector Off	DET Channel # (1–64)	Detector on and off events shall be triggered post any detector delay/extension processing.
82	Detector On	DET Channel # (1–64)	
83	Detector Restored	DET Channel # (1–64)	
84	Detector Fault—Other	DET Channel # (1–64)	Detector failure logged upon local controller diagnostics only (not system diagnostics).
85	Detector Fault—Watchdog Fault	DET Channel # (1–64)	Detector failure logged upon local controller diagnostics only (not system diagnostics).
86	Detector Fault- Open Loop Fault	DET Channel # (1–64)	Detector failure logged upon local controller diagnostics only (not system diagnostics).
87	Detector Fault—Shorted Loop Fault	DET Channel # (1–64)	Detector failure logged upon local controller diagnostics only (not system diagnostics).
88	Detector Fault—Excessive Change Fault	DET Channel # (1–64)	Detector failure logged upon local controller diagnostics only (not system diagnostics).
89	PedDetector Off	DET Channel # (1–16)	Ped detector events shall be triggered post any detector delay/extension processing and may be set multiple times for a single pedestrian call (with future intent to eventually support ped presence and volume).
90	PedDetector On	DET Channel # (1–16)	
91	Pedestrian Detector Failed	Ped Det # (1–16)	Detector failure logged upon local controller diagnostics only (not system diagnostics).
92	Pedestrian Detector Restored	Ped Det # (1–16)	Detector failure logged upon local controller diagnostics only (not system diagnostics).
93–100	Detector events reserved for future use		

TABLE 4.7
Preemption events (23).

Event Code ID	Name	Parameter	Description
101	Preempt Advance Warning Input	Preempt # (1–10)	Set when preemption advance warning input is activated.
102	Preempt (Call) Input On	Preempt # (1–10)	Set when preemption input is activated (prior to preemption delay timing). May be set multiple times if input is intermittent during preemption service.
103	Preempt Gate Down Input Received	Preempt # (1–10)	Set when gate down input is received by the controller (if available).
104	Preempt (Call) Input Off	Preempt # (1–10)	Set when preemption input is de-activated. May be set multiple times if input is intermittent preemption service.
105	Preempt Entry Started	Preempt # (1–10)	Set when preemption delay expires and controller begins transition timing (force-off) to serve preemption.
106	Preemption Begin Track Clearance	Preempt # (1–10)	Set when track clearance phases are green and track clearance timing begins.
107	Preemption Begin Dwell Service	Preempt # (1–10)	Set when preemption dwell or limited service begins or minimum dwell timer is reset due to call drop and reapplication.
108	Preemption Link Active On	Preempt # (1–10)	Set when linked preemptor input is applied from active preemptor.
109	Preemption Link Active Off	Preempt # (1–10)	Set when linked preemptor input is dropped from active preemptor.
110	Preemption Max Presence Exceeded	Preempt # (1–10)	Set when preemption max presence timer is exceeded and preemption input is released from service.
111	Preemption Begin Exit Interval	Preempt # (1–10)	Set when preemption exit interval phases are green and exit timing begins.
112	TSP Check In	TSP #(1–10)	Set when request for priority is received.
113	TSP Adjustment to Early Green	TSP #(1–10)	Set when controller is adjusting active cycle to accommodate early service to TSP phases.
114	TSP Adjustment to Extend Green	TSP #(1–10)	Set when controller is adjusting active cycle to accommodate extended service to TSP phases.
115	TSP Check Out	TSP #(1–10)	Set when request for priority is retracted.
116–130	Preemption events reserved for future use		

TABLE 4.8
Coordination events (23).

Event Code ID	Name	Parameter	Description
131	Coord Pattern Change	Pattern # (0–255)	Coordination pattern that is actively running in the controller. (Highest priority of TOD, System or manual command). This event will not be reapplied if coordination is temporarily suspended for preemption or other external control.
132	Cycle Length Change	Seconds (0–255)	This event shall be populated upon selection of a new coordination pattern change that selects a new cycle length. Cycle lengths in excess of 255 shall record this event with a 255 parameter, requiring controller database lookup for this actual value.
133	Offset Length Change	Seconds (0–255)	This event shall be populated upon selection of a new coordination pattern change that selects a new cycle length. Offsets in excess of 255 shall record this event with a 255 parameter, requiring controller database lookup for this actual value.
134	Split 1 Change	New Split Time in Seconds (0–255)	Split change events shall be populated upon selection of a new coordination pattern as well as during a split change to an active pattern via ACS Lite or other adaptive control system.
135	Split 2 Change		
136	Split 3 Change		
137	Split 4 Change		
138	Split 5 Change		
139	Split 6 Change		
140	Split 7 Change		
141	Split 8 Change		
142	Split 9 Change		
143	Split 10 Change		
144	Split 11 Change		
145	Split 12 Change		
146	Split 13 Change		
147	Split 14 Change		
148	Split 15 Change		
149	Split 16 Change		
150	Coordinated cycle state change	Parameter (0–6) defined as: 0 = Free 1 = In Step 2 = Transition—Add 3 = Transition—Subtract 4 = Transition—Dwell 5 = Local Zero 6 = Begin Pickup	Set when the appropriate coordinator event occurs.
151	Coordinated phase yield point	Phase # (1–16)	Set when the coordinator reaches the yield point for the specified phase.
152–170	Coordination events reserved for future use		

TABLE 4.9
Cabinet/system events (23).

Event Code ID	Name	Parameter	Description
171	Test Input on	Test Input # (as number A = 1, B = 2, etc.)	Cabinet test or special function input as defined by the local controller.
172	Test Input off	Test Input # (as number A = 1, B = 2, etc.)	Cabinet test or special function input as defined by the local controller.
173	Unit Flash Status change	NTCIP Flash state # (0–255)	See NTCIP 1202—2.4.5 for definition (37).
174	Unit Alarm Status 1 change	NTCIP Alarm Status 1# (0–255)	See NTCIP 1202—2.4.8 for definition (37).
175	Alarm Group State Change	NTCIP Alarm Group State (0–255)	See NTCIP 1202—2.4.12.2 for definition (37).
176	Special Function Output on	Special Function # (0–255)	Special function output as defined by the local controller.
177	Special Function Output off	Special Function # (0–255)	Special function output as defined by the local controller.
178	Manual control enable off/on	Manual control enable off/on # (0, 1)	Set when manual control input is applied or removed.
179	Interval Advance off/on	Interval Advance off/on # (0, 1)	Manual signal control input: leading edge on (1), lagging edge (0) optional.
180	Stop Time Input off/on	Stop Time Input Advance off/on # (0, 1)	Set when stop time input is applied or removed, regardless of source.
181	Controller Clock Updated	Optional parameter: Time correction in Seconds (0–255)	Set when the controller OS clock is adjusted via communications, OS command, or external input.
182	Power Failure Detected	True (1)	Line voltage drops between 0 and 89 volts AC for more than 100 ms.
184	Power Restored	True (1)	Line voltage applied/reapplied greater than 98 volts AC.
185	Vendor Specific Alarm	Vendor defined parameter	Placeholder for generic failure/alarm types as defined by vendor.
186–199	Cabinet/System events reserved for future use		

- 171–199. *Cabinet/System Events (Table 4.8)*: indicate miscellaneous controller property–related status changes, such as alarms, clock updates, and power failures.

Events 200–255 are user-defined. Future functions might include logic function status changes and perhaps additional functions for ramp metering signal controllers, as well as control features of adaptive and/or international controllers that do not map into the other existing categories.

For users seeking to implement a data collection system using this schedule of events, it may be desirable to know whether or not a particular event is necessary for developing performance measures. Appendix A provides a table that describes each of the performance measures described in this monograph, with a listing of the required event code IDs needed to generate each one.

4.3 Demonstration

Table 4.10 provides a sample set of discrete events generated by a signal controller with event logging capability, listed in chronological order. The events occurred over a period of 8.5 seconds during the afternoon of October 17, 2012, at Pendleton Pike and Post Road in Indianapolis. Each event has an associated timestamp, event code, and event parameter. The event code describes the type of event that is occurring, and the event parameter indicates the detector or phase number on which the event is being called.

From 13:38:25.6 to 13:38:27.8, four vehicles passed over detectors at the intersection. The first event, code 81, indicates that detector number 43 had just been deactivated owing to a vehicle leaving the detection zone. Subsequently, three other vehicles passed over detectors 28, 30, and 32 in the next 1–2 seconds. At 13:38:28.5, a vehicle having previously occupied detector 26 departs, simultaneously deactivating the presence and the call for phase 2 and indicated by event code 81 and 44, respectively. At 13:38:30.6, event code 7 indicates phase termination on both 2 and 6, as indicated by the event parameter field. Simultaneously, event code 8, begin yellow clearance, was called in the same moment in time for the two phases. At 13:38:30.7, a pedestrian call was put in for phase 2, as indicated by event type 43.

4.4 Case Study

The subsequent material in this monograph demonstrates a series of performance measures for monitoring signal system performance. Data for this study were collected along US 36 (Pendleton Pike) in Indianapolis, Indiana (Figure 4.2). This is an urban arterial carrying commuter traffic between Indianapolis and communities to the northeast. It also serves as an important local road for the town of Lawrence, Indiana, through which it passes. Most of the performance measures in the next two chapters are shown for the intersection of US 36 and Post Road (Figure 4.3).

TABLE 4.10
Excerpt from data table.

Timestamp	Event Code ID	Parameter	Description
2012/10/17 13:38:25.6	81	43	Detector 43 Off
2012/10/17 13:38:26.2	82	30	Detector 30 On
2012/10/17 13:38:26.3	82	32	Detector 32 On
2012/10/17 13:38:26.4	81	32	Detector 32 Off
2012/10/17 13:38:26.4	81	30	Detector 30 Off
2012/10/17 13:38:27.7	82	28	Detector 28 On
2012/10/17 13:38:27.8	81	28	Detector 28 Off
2012/10/17 13:38:28.5	44	2	Phase 2 Call Dropped
2012/10/17 13:38:28.5	81	26	Detector 26 Off
2012/10/17 13:38:30.3	82	17	Detector 17 On
2012/10/17 13:38:30.3	43	6	Phase 6 Call Registered
2012/10/17 13:38:30.6	6	2	Phase 6 Force-off
2012/10/17 13:38:30.6	8	6	Phase 6 Begin Yellow Clearance
2012/10/17 13:38:30.6	6	6	Phase 6 Force-off
2012/10/17 13:38:30.6	8	2	Phase 2 Begin Yellow Clearance
2012/10/17 13:38:30.6	7	6	Phase 6 Green Termination
2012/10/17 13:38:30.6	7	2	Phase 2 Green Termination
2012/10/17 13:38:30.7	2	3	Phase 3 Check
2012/10/17 13:38:30.7	2	4	Phase 4 Check
2012/10/17 13:38:30.7	2	8	Phase 8 Check
2012/10/17 13:38:30.7	43	2	Pedestrian Call Phase 2
2012/10/17 13:38:31.0	82	19	Detector 19 On
2012/10/17 13:38:31.0	82	37	Detector 37 On
2012/10/17 13:38:31.1	81	19	Detector 19 Off
2012/10/17 13:38:31.3	82	39	Detector 39 On
2012/10/17 13:38:31.4	81	39	Detector 39 Off
2012/10/17 13:38:31.7	82	19	Detector 19 On
2012/10/17 13:38:31.8	81	19	Detector 19 Off
2012/10/17 13:38:31.9	82	39	Detector 39 On
2012/10/17 13:38:32.0	81	37	Detector 37 Off
2012/10/17 13:38:32.0	81	39	Detector 39 Off
2012/10/17 13:38:33.4	82	19	Detector 19 On
2012/10/17 13:38:33.6	81	19	Detector 19 Off
2012/10/17 13:38:34.1	81	17	Detector 17 Off

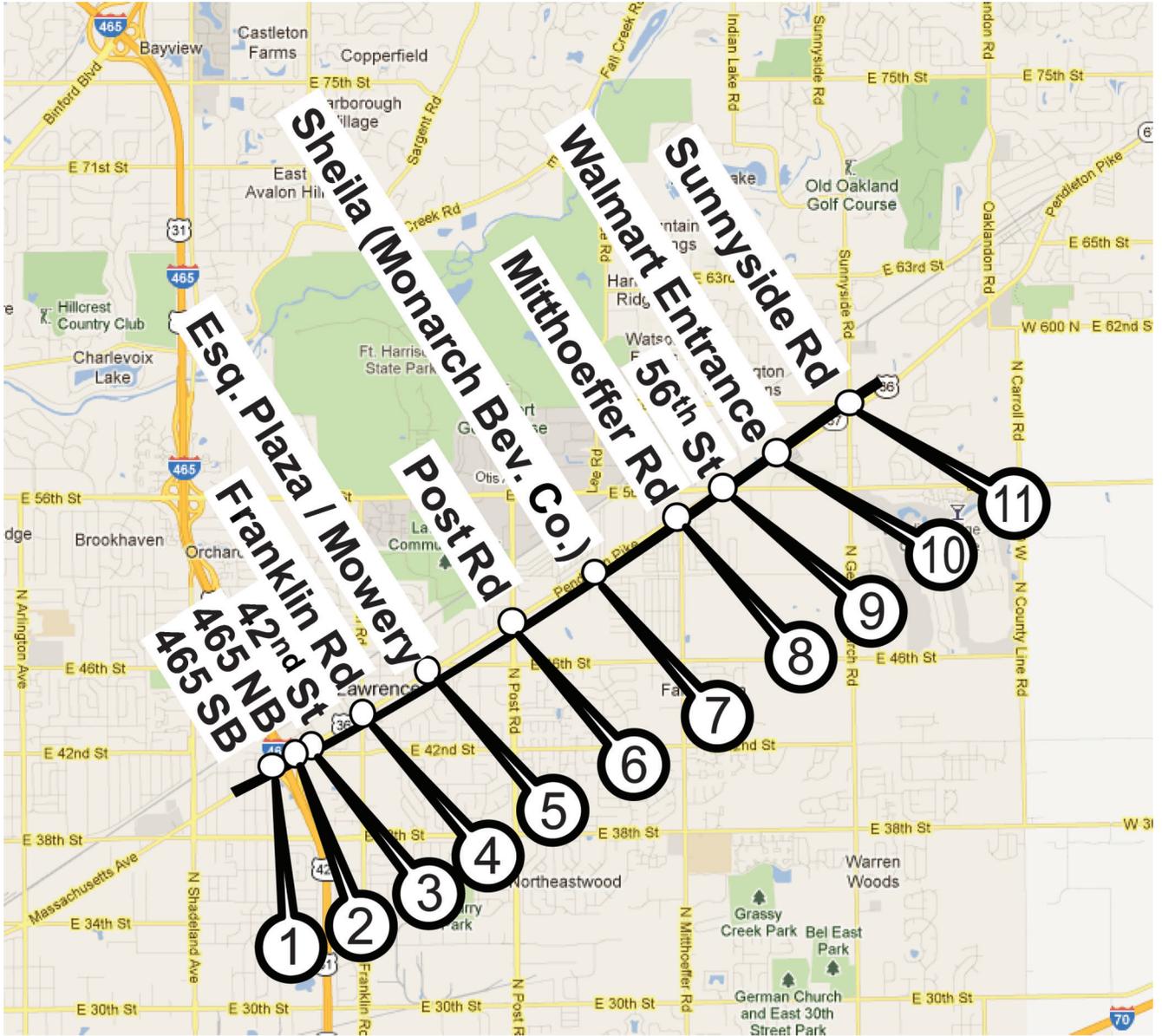


Figure 4.2 US 36 (Pendleton Pike) in Indianapolis, Indiana.

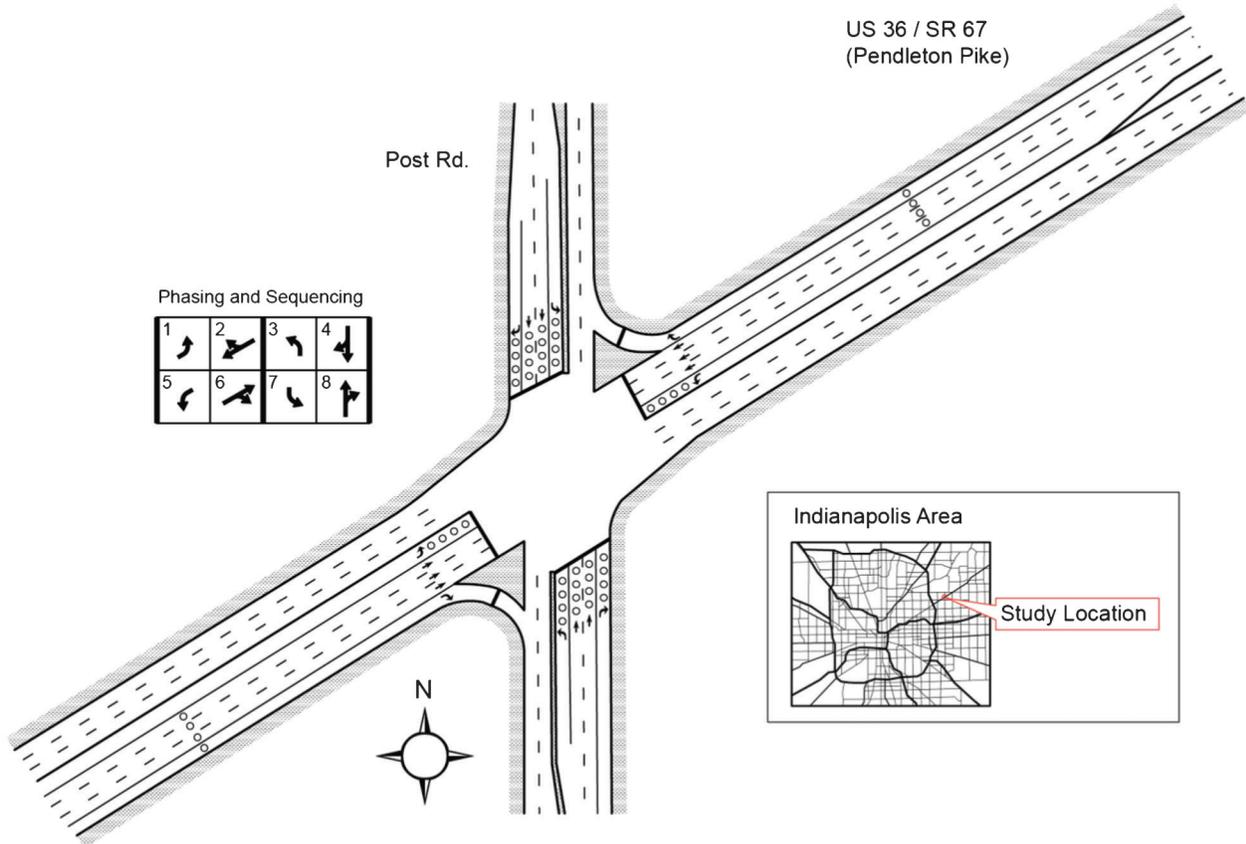


Figure 4.3 Pendleton Pike and Post Road.

5. CAPACITY PERFORMANCE MEASURES

This chapter demonstrates performance measures for monitoring the utilization of capacity at a signalized intersection. These include measures of the allocation of signal time and of the number of vehicles served by each phase.

5.1 Traffic Volume Visualization

Traffic volumes support a number of purposes in the management of traffic control systems. They are particularly important for determining traffic signal timings and for planning purposes. This discussion of performance measures opens with alternative views of the traffic volume levels using 15-minute counts converted to the equivalent hourly flow rate. This type of data can be obtained from a variety of sources besides high-resolution data.

Figure 5.1 shows the traffic volume for each lane (except for the right-turn lanes, which do not have counting detectors) at US 36 and Post Road on October 17, 2012. The plots reveal that US 36 has the dominant volumes and very characteristic a.m. and p.m. peaks in the westbound and eastbound through lanes, respectively. There are some other trends in the other lane volumes, but they are less drastic. Figure 5.2

and Figure 5.3 show the volume aggregated from lane to phase, in units of vehicles per hour (Figure 5.2) and vehicles per hour per lane (Figure 5.3). The eight plots are ordered according to the local phase sequence, which is the standard eight-phase dual-ring configuration. The same trends in the time-of-day peaks can be observed in these plots, somewhat more compactly. The rest of this chapter focuses on the eight-phase performance measure view, as it reveals the operation of each individual phase within one concise graphic.

This type of data could easily be entered into a variety of applications, such as signal timing optimization software or HCM-based or similar analysis tools. The data are also relevant to planning applications. Average daily traffic (ADT) values can be obtained by summing across lanes on an approach. INDOT has used automatically collected vehicle counts for similar purposes for over 10 years, using counts obtained from count detector cards that were automatically logged in a traffic count database.

The next two figures show another view of traffic data by showing the relative flow rates on the inbound and outbound movements along US 36 during the a.m. (Figure 5.4) and p.m. (Figure 5.5) peak hours. In these plots, the width of the line indicates the volume of traffic; the different colors indicate the two directions (inbound and outbound) along the arterial. As

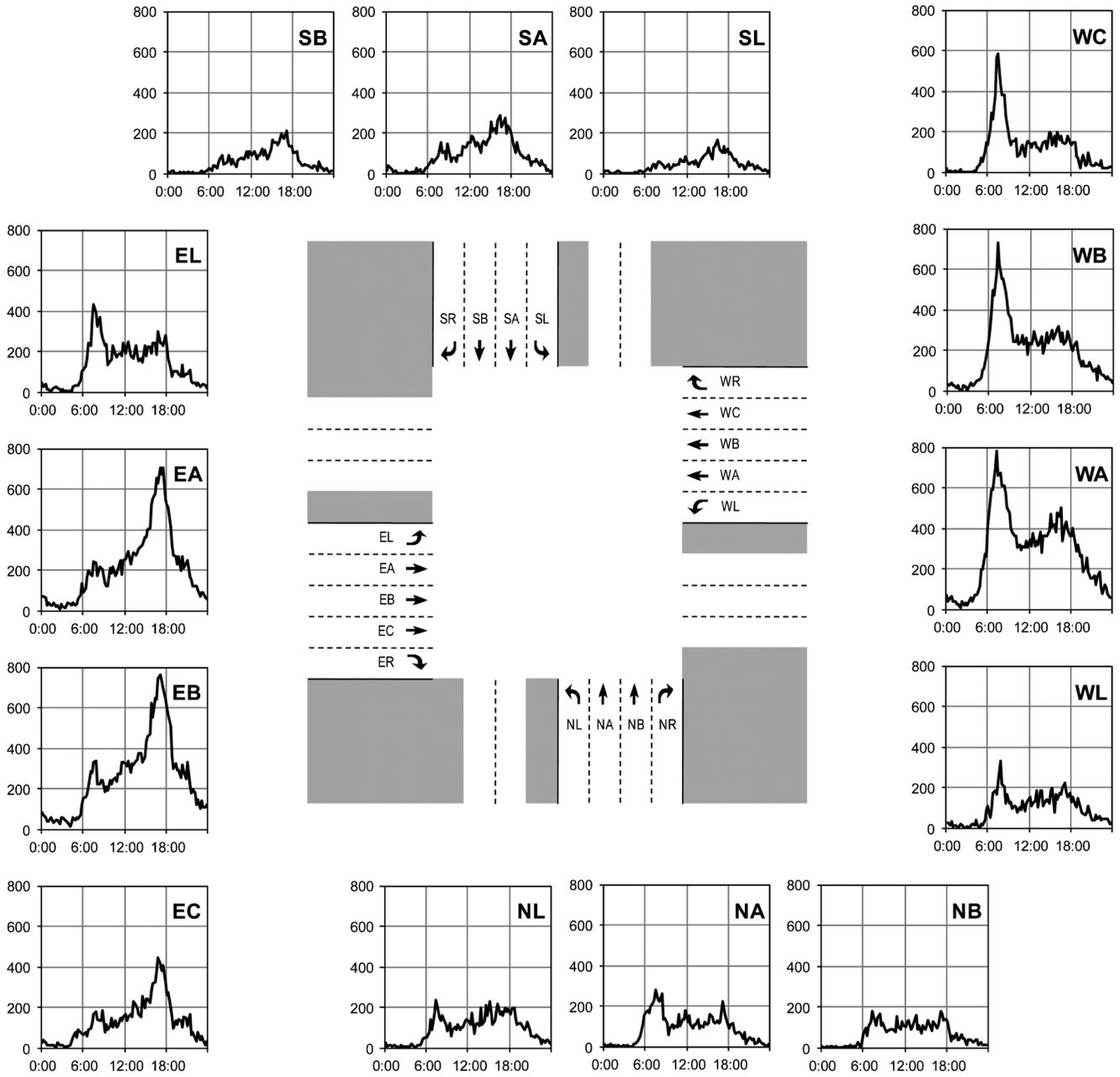


Figure 5.1 15-minute vehicle volumes by lane.

expected, inbound volumes dominate during the a.m. (Figure 5.4) and outbound volumes dominate during the p.m. (Figure 5.5). Although this information was likely known beforehand, a few other details are observable. In Figure 5.4, it is apparent that inbound traffic is relatively light east of Sunnyside Road, suggesting that a substantial portion of the traffic east of that intersection originates from side-street entries at the intersection with Sunnyside Road. Also, in Figure 5.5, the p.m. flows between the interstate ramps and Franklin Road are somewhat balanced between the inbound and outbound directions. The outbound

traffic is strongest on the sections to the east of Franklin Road. This suggests that Franklin Road is the source of a considerable amount of outbound traffic.

These data views are based on traffic counts alone, with no reference to the signal data. Although these are useful to obtain a broad view of the overall traffic demand, they reveal little about the quality of operations. For this purpose, it is useful to also track the phase timing to be able to quantify how well capacity is being apportioned among phases. The remainder of this chapter is dedicated to performance measures for that purpose.

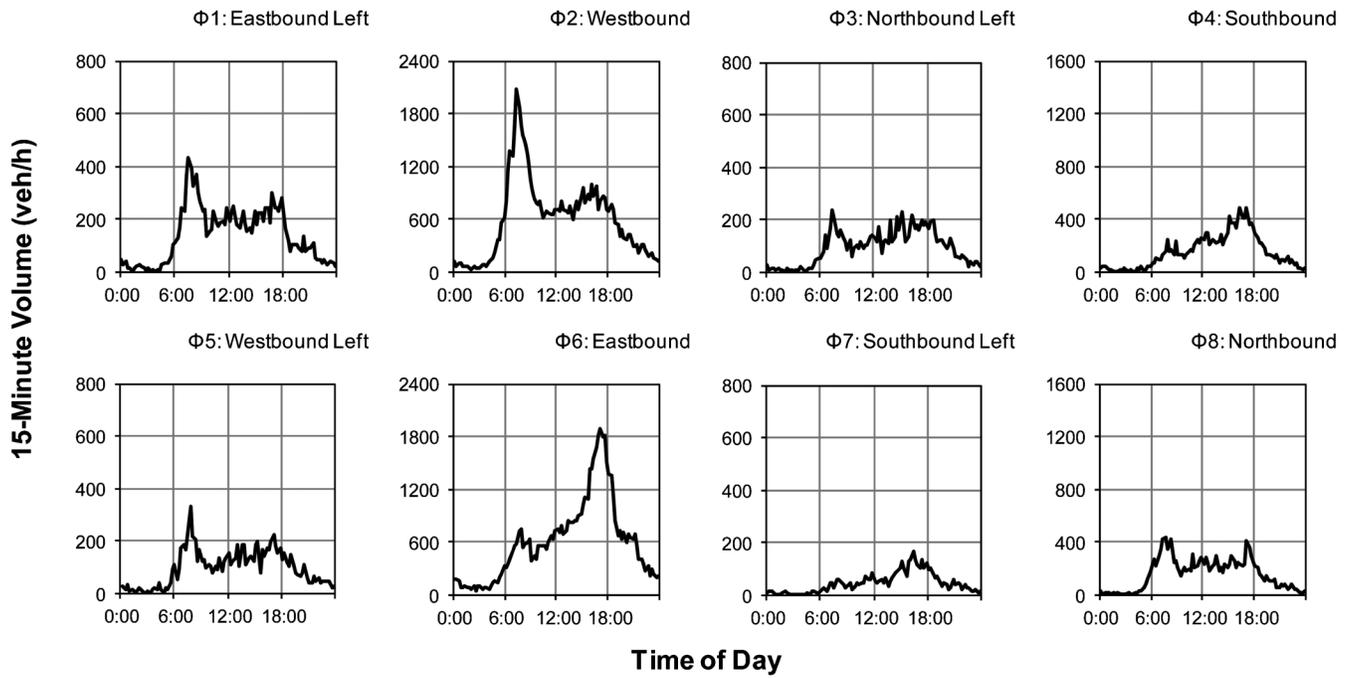


Figure 5.2 Vehicle volume by phase (vehicles per hour).

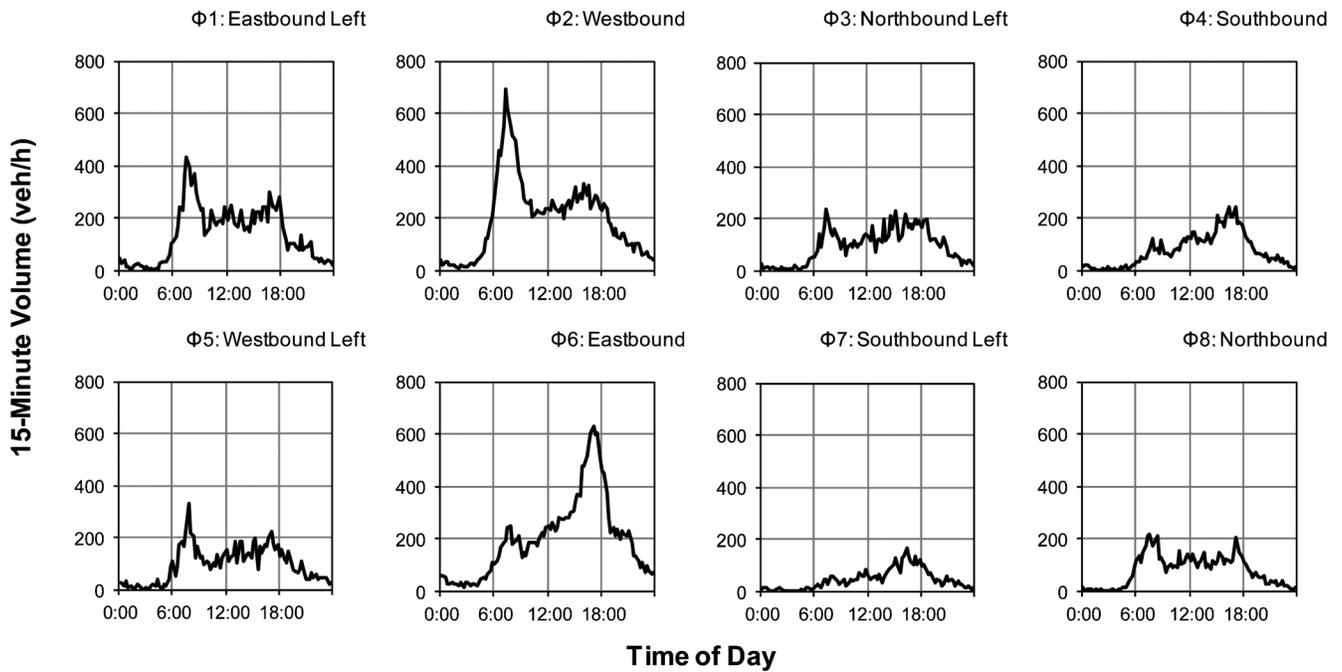


Figure 5.3 Vehicle volume by phase (vehicles per hour per lane).

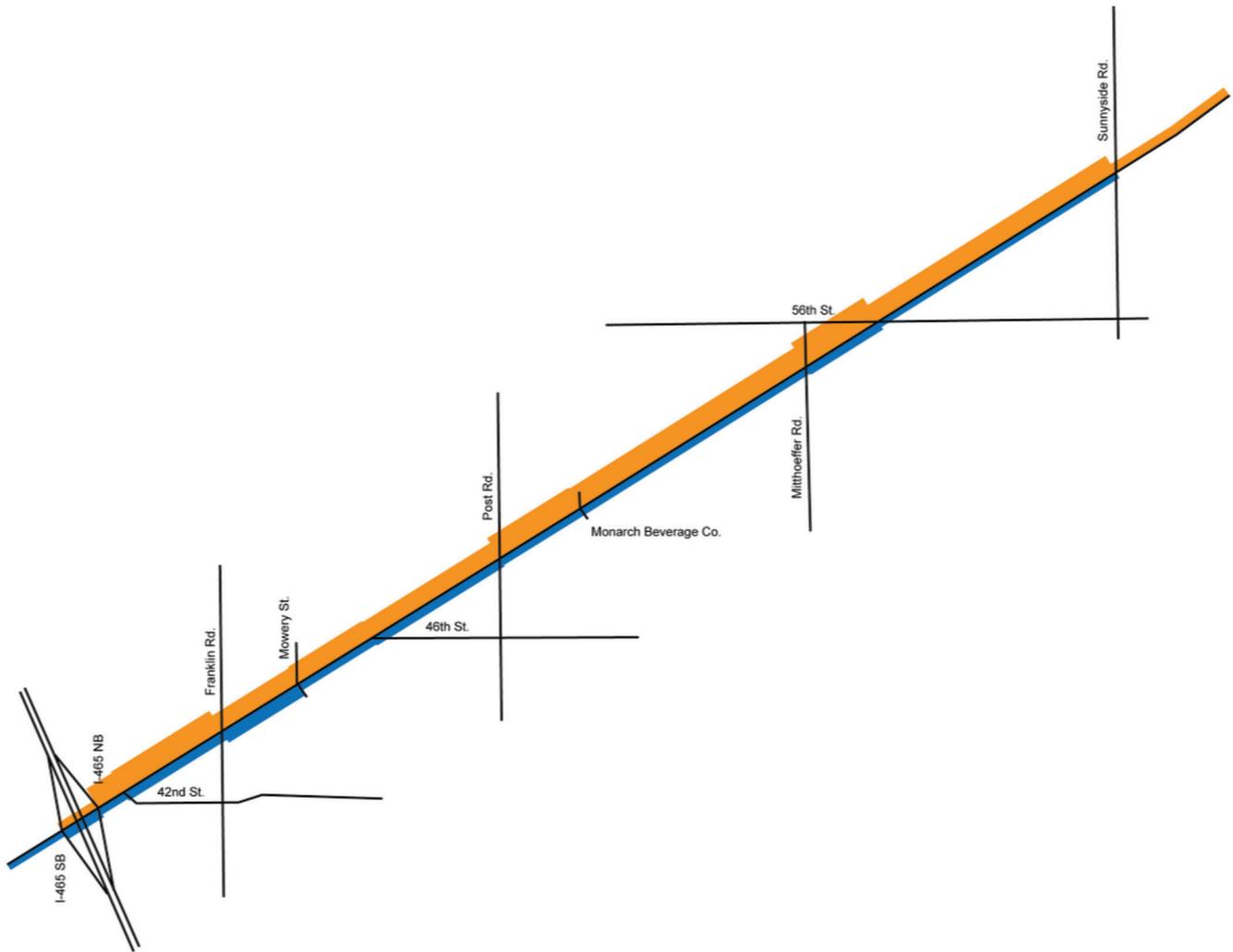


Figure 5.4 Corridor peak hour volumes (a.m.).

5.2 Capacity Utilization Concepts

To begin examining the effectiveness of capacity allocation by the signal controller, it is necessary to first discuss some capacity concepts. It is helpful to begin by defining traffic flow: the volume v of traffic passing a point during a unit of time:

$$v = \frac{N}{T} \tag{5.1}$$

Here, N is the vehicle count during an arbitrary time period T . This is often expressed as the number of vehicles per hour, although the actual counting interval can be much smaller. Although “flow rate” suggests a continuous variable that changes over time, ultimately it depends on discrete events, the individual vehicle counts. The time between two successive vehicle counts is referred to as the headway, h . The total analysis time

may also be considered as a sequence of successive headways:

$$T = \sum_{i=1}^N h_i \tag{5.2}$$

The flow rate and the *average* headway are related to each other as follows (38):

$$v = \frac{N}{T} = \frac{N}{\sum_i h_i} = \frac{1}{\frac{1}{N} \sum_i h_i} = \frac{1}{\bar{h}} \tag{5.3}$$

If N is the number of vehicles arriving at an intersection during a given *signal cycle*, where C is the cycle length in seconds, then the *equivalent hourly volume* is given by

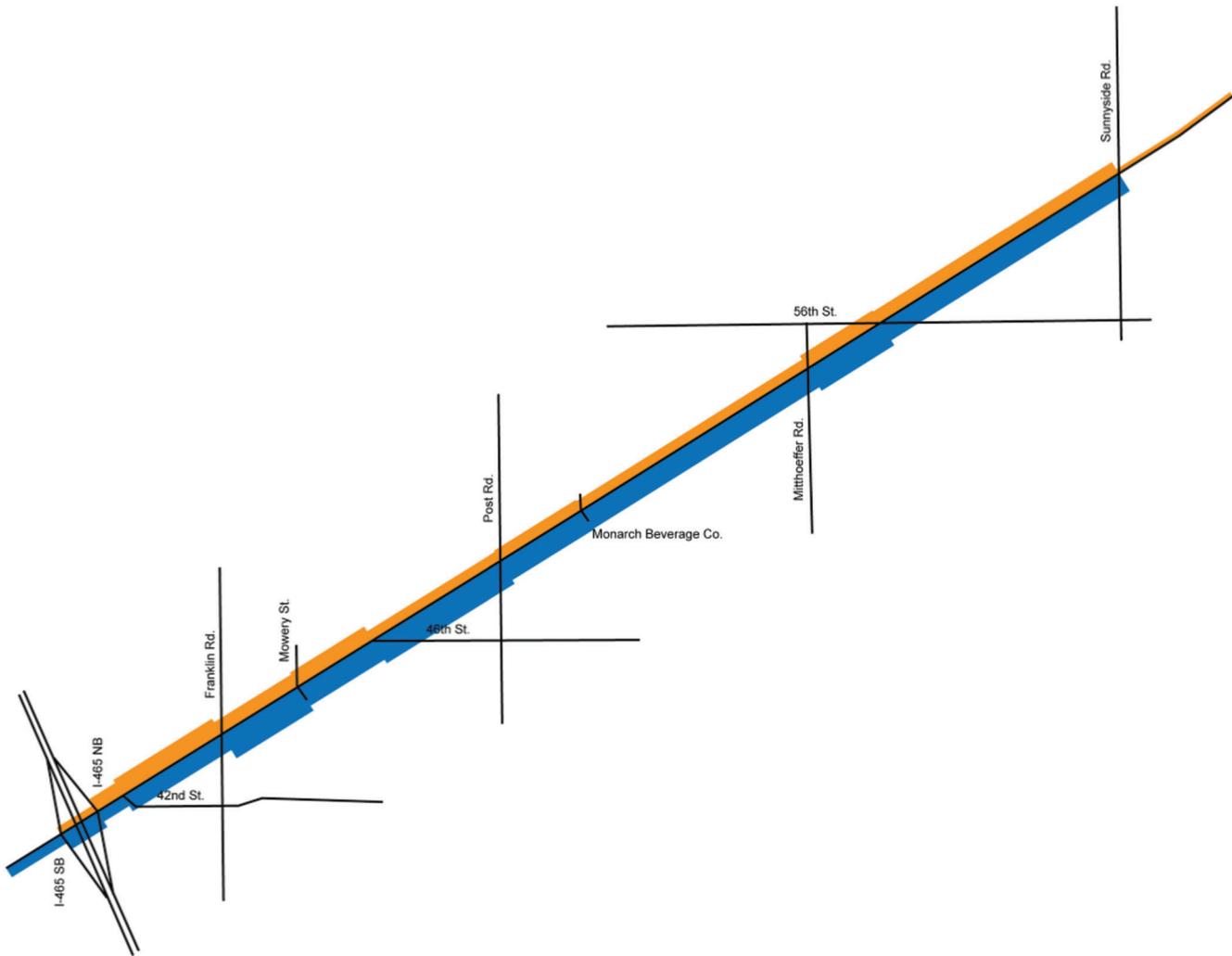


Figure 5.5 Corridor peak hour volumes (p.m.).

$$v = N \frac{3600}{C} \tag{5.4}$$

$$c_H = c \frac{3600}{C} = s_H \frac{g}{C} \tag{5.6}$$

The *saturation flow rate* at a location is defined as the theoretical maximum possible flow rate on a facility; from this, the *capacity* of a roadway is defined as the theoretical maximum possible number of vehicles that can move through that roadway during a given time interval. For a traffic signal, the capacity c of a particular movement in a given cycle can be calculated, in units of vehicles (39), as

$$c = sg \tag{5.5}$$

where s is the saturation flow rate in vehicles per second and g is the green time in seconds. The *equivalent hourly capacity* of the capacity provided in a given signal cycle can be computed by the formula

where s_H is the saturation flow rate in vehicles per hour, g is the green time in seconds, and C is the cycle length in seconds. In the remainder of this chapter, c and c_H are used interchangeably, with the understanding that the values are equivalent but the units are changed. To understand this conversion, it is helpful to remember that c has units of vehicles within one cycle. Dividing by the cycle length allows a conversion to vehicles per hour. However, on a more fundamental level, capacity during any particular cycle is a function only of the green time given and is independent of the cycle length.

The concept of *effective green time* is illustrated in Figure 5.6, which shows a typical plot of traffic flow exiting from a signalized link. The effective green time begins slightly later than the actual green indication

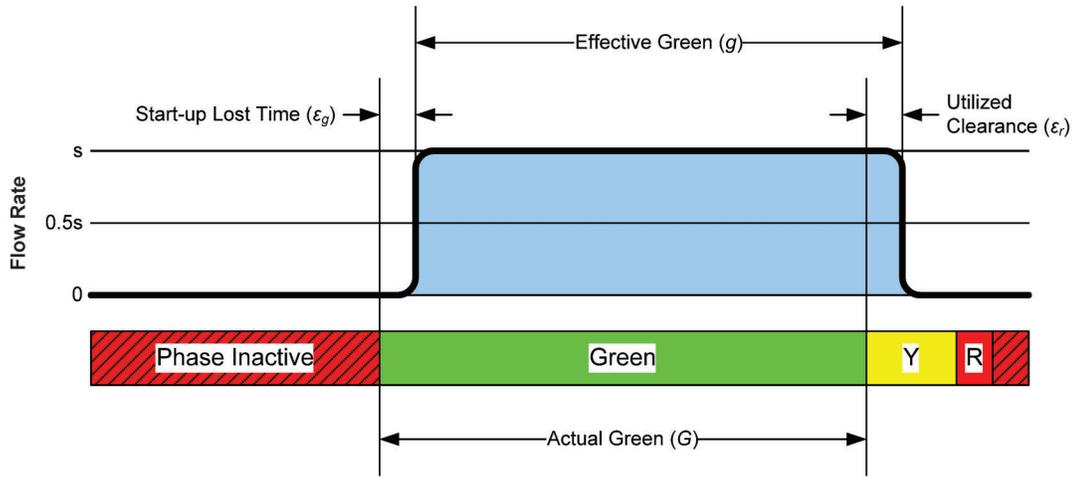


Figure 5.6 Theoretical utilization of capacity by a fully saturated movement.

because of the *start-up lost time* associated with driver reaction to the changing signal. It also ends slightly later because of *driver utilization of the clearance period*, attributable not only to reaction time but also to driver attitude toward yellow, which varies by location and with traffic conditions. In this report, since they adjust the beginning of green and beginning of red, we define these quantities respectively as ϵ_g and ϵ_r . These are applied as follows:

$$t_g = t_G + \epsilon_g, \quad (5.7)$$

$$t_r = t_Y + \epsilon_r, \quad (5.8)$$

$$g = G - \epsilon_g + \epsilon_r. \quad (5.9)$$

In this system of equations,

t_g = time when effective green begins;

t_G = time when actual green indication begins;

ϵ_g = start-up lost time;

t_r = time when effective red begins;

t_Y = time when actual green indication ends (beginning of yellow);

ϵ_r = amount of clearance time utilized for vehicle movement;

g = duration of effective green (as above);

G = duration of the actual green indication.

From the definition of capacity in Equation 5.5, the *degree of saturation*, or *volume-to-capacity ratio* x , is defined as

$$x = \frac{v}{c} = \frac{v}{s} \frac{s}{C} = \frac{vC}{sg}, \quad (5.10)$$

where the terms are as defined earlier. Note that the following definition is also valid if we substitute Equation 5.4 into Equation 5.10, demonstrating that

x is dependent only on g , and not explicitly on C :

$$x = \frac{3600 N}{s g}. \quad (5.11)$$

In measuring capacity, the analyst must select values for ϵ_g , ϵ_r , and s . The Highway Capacity Manual (HCM) (40), for example, recommends using values of 2 seconds for both ϵ_g and ϵ_r , and it provides a methodology for calculating s based on a number of factors such as lane geometry and percentage of heavy vehicles. We will use 1900 vehicles per hour per lane. In practice, however, there are often locations where actual flow rates exceed this value, thus leading to situations where $v > c$ and consequently $x > 1.0$. Therefore, c should not be thought of as the actual maximum possible flow rate, but is actually more of an expected capacity based on the assumed value of s , which might be considered a “design” saturation flow rate.

A 2004 JTRP study (41) included field observations of saturation flow rates, start-up lost time, and driver utilization of clearance from various intersections from several locations throughout the state of Indiana. Some findings from this study are summarized below.

- A summary of the field observations of those parameters are outlined in Table 5.1, indicating considerable variation in the values.
- Saturation flow rates in the city of Indianapolis were found to be up to 423 vehicles per hour per lane greater than saturation flows in small towns.
- Recommended saturation flow rates for use in the state of Indiana are shown in Table 5.2.
- The HCM recommended values for ϵ_g and ϵ_r were found to be appropriate for use in Indiana.

5.3 Cycle Length

Perhaps the most basic cycle-by-cycle performance measure is the duration of the cycle length. Although this value is often known ahead of time, it is still a good

TABLE 5.1
Capacity parameters from field observation, from Perez-Cartagena and Tarko (41).

Parameter		Average	St. Dev.	Minimum	Maximum
Through	e_g	1.87	0.74	0.53	3.91
	s	1842	199	1352	2178
Left	e_g	1.61	0.71	0.57	2.91
	s	1844	117	1764	2079
Left and Through	e_r	2.81	1.26	0.03	5.83

idea to confirm that the correct *background* cycle length was programmed into the controller, and it is also useful to measure the *effective* cycle length that is produced by the controller as a consequence of phase actuation. Measuring from signal indications gives a measure of the effective cycle length, whereas measuring from an event such as the yield point or local zero gives the background cycle length.

Figure 5.7 shows a plot of the background cycle length measured from the yield points of phase 2 occurring on October 17, 2012, at US 36 and Post Road. This intersection runs under a coordinated plan between 6:00 and 22:00, as indicated in this plot. A cycle length of 100 seconds is used throughout the entire day. Table 5.3 shows a sample of the data where the cycle length is measured as the time between two successive yield points. The three spikes in Figure 5.7 occur when subsequent yield points were longer than 100 seconds, which occurs during controller transitioning between different signal patterns. Yield points are not recorded during transition. Four different patterns are used throughout the day that implement different splits and offsets for time of day, while the cycle length remains the same.

The *effective* cycle length is shown in Figure 5.8. Here, cycle length is measured by finding the barrier crossing between phase groups {1, 2, 5, 6} and {3, 4, 7, 8} (see Figure 3.3). This definition of cycle length is less sensitive to changes in the lead-lag configuration of phases within each group, and fluctuations due to actuation. Between 6:00 and 22:00, the cycle length fluctuates around 100 seconds. The variations in the measurement are caused by actuation of the coordinated phase (see Figure 3.17), which allows the *effective* cycles to be shorter or longer than others by several seconds. The early morning (0:00–6:00) and late evening (22:00–24:00) time periods experience a great deal of fluctuation in the effective cycle length. During

TABLE 5.2
Recommended base saturation flow rates (vehicles per hour per lane), from Perez-Cartagena and Tarko (41).

Number of Lanes	Population		
	<20,000	20,000–100,000	Indianapolis Area
1	1540	1800	1960
2	1580	1840	2010
3	1600	1860	2020

these time periods, the intersection operates in fully actuated mode, so there is no fixed cycle length, and the time needed to serve all phases fluctuates considerably, from as low as 38 seconds (where phases are served their minimum times and many are omitted) to hundreds of seconds (where there is no side street demand, and the controller dwells in phases 2/6).

The computation of cycle length from the listing of phase events is shown in Table 5.4 and Table 5.5. Table 5.4 presents an ordered list of the beginning of green times for all eight phases at the intersection, as indicated by the changes in the phase states in each row. The current active phase group is shown as are the boundaries of cycles marked by a transition from group 1 to group 2. The final column shows the current cycle number. New cycles begin at 13:30:17.0, 13:31:57.0, 13:33:36.6, and 13:35:17.0. There is considerable variation in the included phases from one cycle to the next; the 513th cycle in particular includes only one left-turn phase (phase 7). The cycle boundaries are shown by themselves in Table 5.5, along with the effective cycle length computed from the time between successive boundaries. The pattern of a short cycle followed by a long cycle is typical of operations with actuated coordinated phases. For example, cycle $i = 513$ ends 1.4 seconds early because of termination of the coordinated phases; this time is absorbed by the 514th cycle, which is 101.4 seconds. There is a similar trade of time between the 515th and 516th cycles. Figure 5.8 shows that this occurs rather frequently throughout the day, except during the a.m. and p.m. peaks, when the coordinated phases usually extend past the yield point and retain the entire programmed split.

Figure 5.9 shows a useful application of these data in the validation of signal operations. This chart includes two lines corresponding to the cycle lengths at two neighboring intersections at a different location. The pattern change times of Intersection 2 occur one hour earlier than those of Intersection 1; the presence of such an error suggests that the pattern start times, time zone, or daylight savings time settings might have been incorrectly programmed. Once corrected, we would expect the plots to show the transition spikes occurring at approximately the same time.

Another example application of cycle length data is illustrated by Figure 5.10. This shows the trace of cycle length at an intersection under five different control strategies, under identical traffic conditions

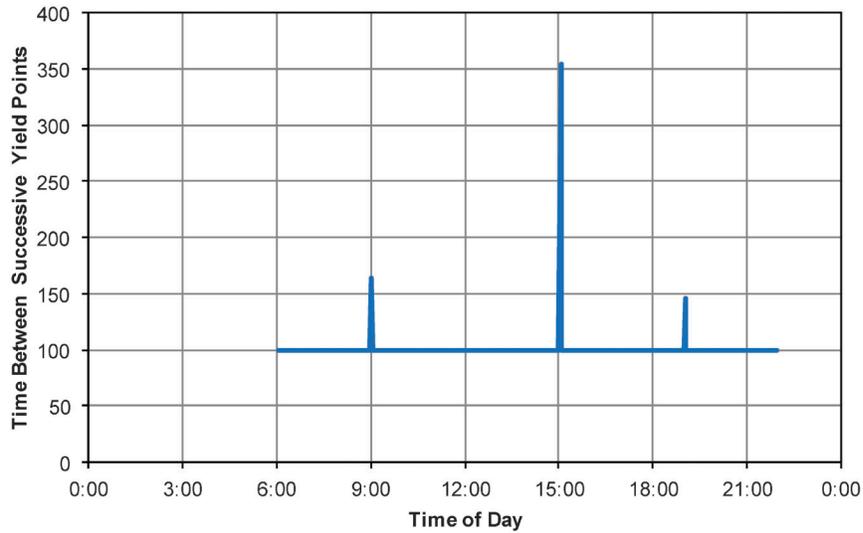


Figure 5.7 Background cycle length.

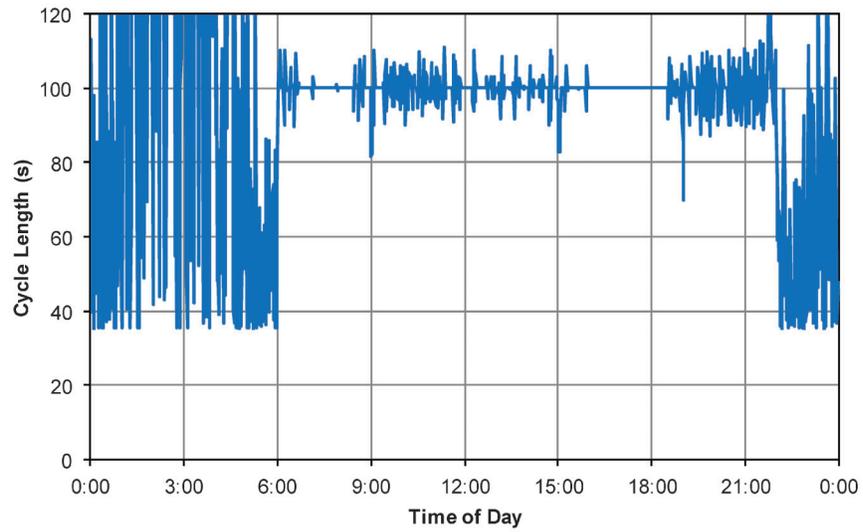


Figure 5.8 Effective cycle length.

TABLE 5.3
Yield point cycle length calculation.

Date	Yield Point Time	Time to Next Yield Point	Cycle Length (s)
10/17/2012	13:23:20.6	00:01:40.0	100.0
10/17/2012	13:25:00.6	00:01:40.0	100.0
10/17/2012	13:26:40.6	00:01:40.0	100.0
10/17/2012	13:28:20.6	00:01:40.0	100.0
10/17/2012	13:30:00.6	00:01:40.0	100.0
10/17/2012	13:31:40.6	00:01:40.0	100.0
10/17/2012	13:33:20.6	00:01:40.0	100.0
10/17/2012	13:35:00.6	00:01:40.0	100.0
10/17/2012	13:36:40.6	00:01:40.0	100.0
10/17/2012	13:38:20.6	00:01:40.0	100.0
10/17/2012	13:40:00.6	00:01:40.0	100.0

TABLE 5.4
Real-time cycle length calculation.

Date	Event Time		Phase States								PhaseGroup	Cycle	
	Timestamp	Seconds	1	2	3	4	5	6	7	8		Boundary	<i>i</i>
10/17/2012	13:29:21.0	48561.0	0	1	0	0	1	0	0	0	1	0	511
10/17/2012	13:29:36.5	48576.5	0	1	0	0	0	1	0	0	1	0	511
10/17/2012	13:30:17.0	48617.0	0	0	1	0	0	0	0	1	2	1	512
10/17/2012	13:30:31.0	48631.0	0	0	0	1	0	0	0	1	2	0	512
10/17/2012	13:30:52.3	48652.3	1	0	0	0	1	0	0	0	1	0	512
10/17/2012	13:31:07.6	48666.6	0	1	0	0	1	0	0	0	1	0	512
10/17/2012	13:31:07.5	48667.5	0	1	0	0	0	1	0	0	1	0	512
10/17/2012	13:31:57.0	48717.0	0	0	0	1	0	0	1	0	2	1	513
10/17/2012	13:32:11.0	48731.0	0	0	0	1	0	0	0	1	2	0	513
10/17/2012	13:32:27.9	48746.9	0	1	0	0	0	1	0	0	1	0	513
10/17/2012	13:33:36.6	48815.6	0	0	1	0	0	0	0	1	2	1	514
10/17/2012	13:33:51.0	48831.0	0	0	0	1	0	0	0	1	2	0	514
10/17/2012	13:34:07.3	48847.3	0	1	0	0	1	0	0	0	1	0	514
10/17/2012	13:34:20.2	48860.2	0	1	0	0	0	1	0	0	1	0	514
10/17/2012	13:35:17.0	48917.0	0	0	1	0	0	0	1	0	2	1	515
10/17/2012	13:35:31.0	48931.0	0	0	0	1	0	0	0	1	2	0	515

using software-in-the-loop simulation (42). The five traces show the cycle length during free conditions (FREE), conventional time-of-day operation (TOD), time-of-day with adaptive control (TOD+ACS), traffic responsive pattern selection (TR), and traffic responsive with adaptive control TR+ACS). Although this information would need to be combined with more detailed evaluation of capacity and progression performance to make a complete appraisal of the five different strategies, the cycle length reveals that free operation takes the least amount of time to serve all of the demand in a given cycle (of course, without any provision for coordination), whereas the other strategies impose cycle lengths of different values, and adaptive control causes occasional short and/or long cycles, about once or twice per hour, when the timing parameters are adjusted.

5.4 Green Time and Capacity Allocation

The next interval measurement after the cycle time is of the individual phase green times. These can be measured similarly from the event data. As discussed earlier, the *effective* green time differs from the actual green time because of the fact that vehicle movement begins a few seconds after the beginning of green, and a portion of the clearance interval is utilized for vehicle movement (see Figure 5.6).

Table 5.6 presents a list of service instances of phase 2 at US 36 and Post Road from the afternoon of October 17, 2012. Each instance is defined by the series of green and yellow times found in the source data. These two timestamps are adjusted by the start-up lost time ϵ_g (Equation 5.7) and the amount of utilized clearance ϵ_r (Equation 5.8) to yield the beginning and

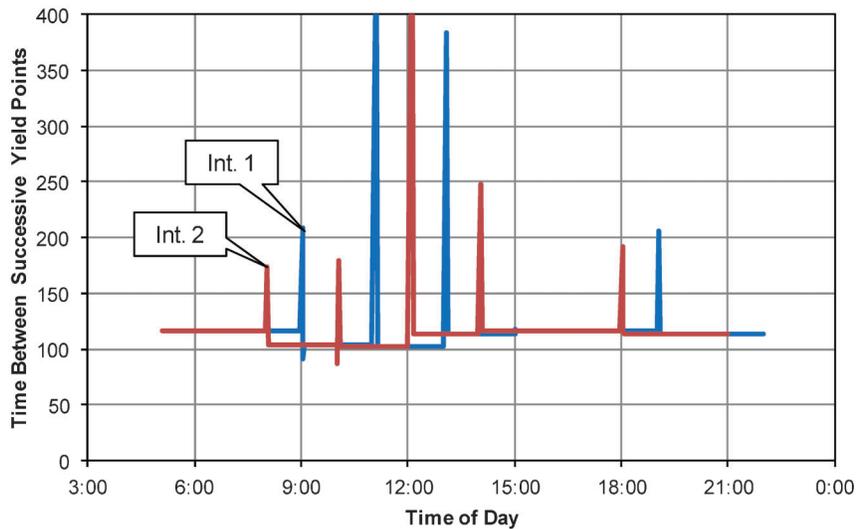


Figure 5.9 Plot of cycle length at two intersections where pattern transitions at one intersection occur one hour earlier (hypothetical example of a time zone error).

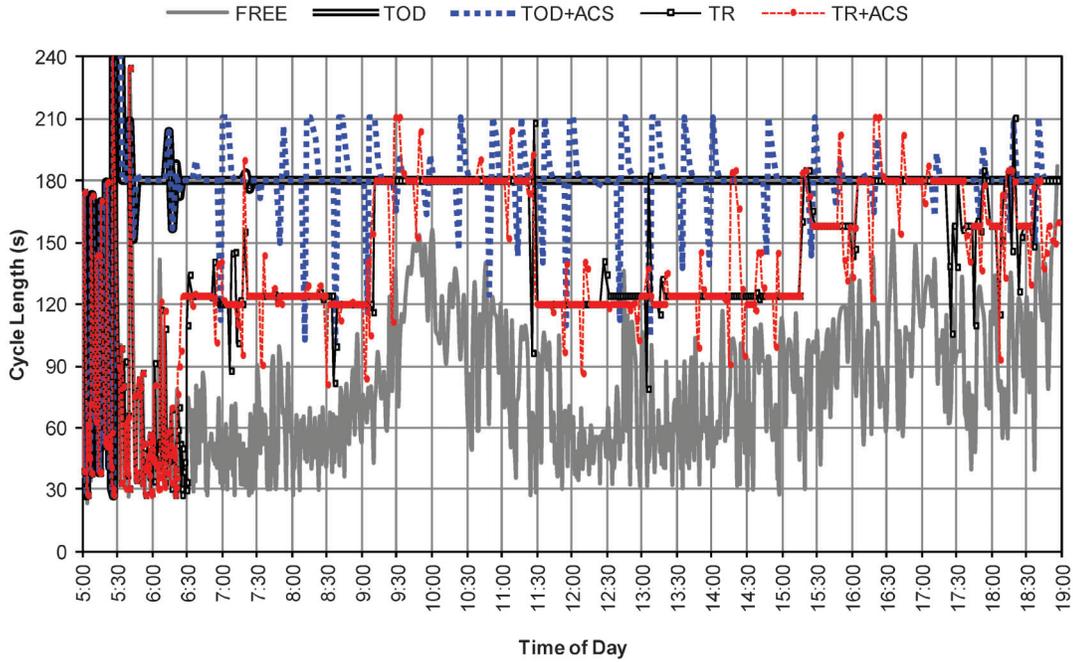


Figure 5.10 Cycle length under different control types (42).

end of the effective green. The green time for each interval is obtained from the time difference between those events. The equivalent capacity (in vehicles per hour) is then found by multiplying by the saturation flow rate (Equation 5.5). In Table 5.6, the saturation flow rate s does not vary; however, improvements in detection (43) or a more accurate model of s (44) might lead to refinements allowing s to be determined for individual service instances.

The next step is to aggregate the individual times from individual service instances to cycles. Table 5.7 lists the cycles that the service instances of Table 5.6 belong to. Because the example data look at a coordinated phase in a conventionally phased intersection, there is one service instance of the phase in each cycle throughout the day, and $j = i$. Because of actuation, phase re-service, and so forth, there is a one-to-many relationship between the cycles and service instances. When multiple service instances occur, their green times and capacities are summed to provide a total green time g_i and a total capacity c_i for the i th

TABLE 5.5
Cycle length data table.

Date	$t_{BOC,i}$		C_i	i
	Timestamp	Seconds		
10/17/2012	13:30:17.0	48617.0	100.0	512
10/17/2012	13:31:57.0	48717.0	98.6	513
10/17/2012	13:33:36.3	48815.6	101.4	514
10/17/2012	13:35:17.0	48917.0	91.6	515
10/17/2012	13:36:49.6	49008.6	108.4	516
10/17/2012	13:38:37.0	49117.0	100.0	517

cycle. The overall portion of cycle time used to serve the phase is found from the green-to-cycle ratio g_i/C_i .

The calculation of capacity for protected-permitted movements requires that the *utilization* of other phases be known (or at least estimated) beforehand. The calculation details are discussed in Section 5.6.

The next series of figures show plots of g , c , and g/C , respectively, for phase 2 in Figure 5.11, Figure 5.13, and Figure 5.15; and for all eight phases in Figure 5.12, Figure 5.14, and Figure 5.16. The lines in these plots represent a 10-cycle moving average that highlights the trend in the data throughout the day. Generally, a longer green time corresponds to a higher amount of capacity and a greater share of the total cycle. One notable difference in the plots is that the long green times seen during overnight periods in Figure 5.11 are normalized into the equivalent hourly capacity in Figure 5.13. Similar effects can be observed in the other phases in comparing Figure 5.12 with Figure 5.14. The impact of peak hour volumes can be seen in Figure 5.11 by the flat-lining of the phase 2 green time during the peak periods (i.e., around 7:00 and between 15:00 and 18:00), a consequence of phases attaining the longest possible green times because of high demand during the peaks.

In the eight-phase plots of green time (Figure 5.12) and g/c (Figure 5.16), it is clear that phases 2 and 6 obtain the largest share of green time, with phases 4 and 8 being the next most dominant, and the left-turn phases having the least share of the green time. Note that this represents the *protected* green times. Those phases that control left turns that also have a permitted component will acquire additional capacity from those permitted periods. Thus, the eight-phase plots of capacity (Figure 5.14) show higher capacities for the

TABLE 5.6
Calculation of effective red and green (phase 2).

Date	φ	j	$t_{G,2,j}$	$t_{Y,2,j}$	ε_g	ε_r	$t_{g,2,j}$	$t_{r,2,j}$	$g_{2,j}$	s_2	$c_{2,j}$
10/17/2012	2	512	13:31:06.6	13:31:50.6	2.0	2.0	13:31:08.6	13:31:52.6	44.0	5700	69.7
10/17/2012	2	513	13:32:26.9	13:33:29.2	2.0	2.0	13:32:28.9	13:33:31.2	62.3	5700	98.6
10/17/2012	2	514	13:34:07.3	13:35:10.6	2.0	2.0	13:34:09.3	13:35:12.6	63.3	5700	100.2
10/17/2012	2	515	13:36:01.2	13:36:42.2	2.0	2.0	13:36:03.2	13:36:44.2	41.0	5700	64.9
10/17/2012	2	516	13:37:42.9	13:38:30.6	2.0	2.0	13:37:44.9	13:38:32.6	47.7	5700	75.5
10/17/2012	2	517	13:39:07.8	13:40:10.6	2.0	2.0	13:39:09.8	13:40:12.6	62.8	5700	99.4

TABLE 5.7
Green time data table (phase 2).

Date	$t_{BOC,i}$	i	C_i	φ	j	$t_{G,2,j}$	$g_{2,j}$	$c_{2,j}$	$g_{2,i}$	$c_{2,i}$	$g_{2,i}/C_i$
10/17/2012	13:30:17.0	512	100.0	2	512	13:31:06.6	44.0	69.7	44.0	69.7	0.44
10/17/2012	13:31:57.0	513	98.6	2	513	13:32:26.9	62.3	98.6	62.3	98.6	0.63
10/17/2012	13:33:36.3	514	101.4	2	514	13:34:07.3	63.3	100.2	63.3	100.2	0.62
10/17/2012	13:35:17.0	515	91.6	2	515	13:36:01.2	41.0	64.9	41.0	64.9	0.45
10/17/2012	13:36:49.6	516	108.4	2	516	13:37:42.9	47.7	75.5	47.7	75.5	0.44
10/17/2012	13:38:37.0	517	100.0	2	517	13:39:07.8	62.8	99.4	62.8	99.4	0.63

left-turn phases than are implied by protected green time alone (Figure 5.12). This is particularly evident for phases 1 and 5 during the early morning periods when the signal dwells in phases 2 and 6, while the volumes are low, meaning that the left-turn phases are provided an enormous measure of permitted capacity during each long effective cycle. Details on the calculation of permitted phase capacities are discussed in Section 5.6.

5.5 Volume and Capacity Utilization

The capacity, green time, and phase termination performance measures shown in the previous sections can be applied at any intersection regardless of its detection capabilities. When vehicle counts are available, a more detailed portrait of the intersection

performance can be generated that reveals the degree to which each phase is utilized by traffic, as well as the intersection as a whole. One important detail that should preface this discussion is the fact that automatic vehicle counts are a measure of the *served* volume, not the *demand*. The only vehicles that are counted are those that can cross the detector to be counted. This should be taken into account when examining vehicle count based data for congested periods where the demand exceeds the capacity.

An example of raw count data is shown in Table 5.8. This shows individual vehicle count times measured from the advance detector for phase 2. The lane where each vehicle was counted is shown in the second column. The next column shows the cumulative vehicle count, shown with the index k . The detection time of

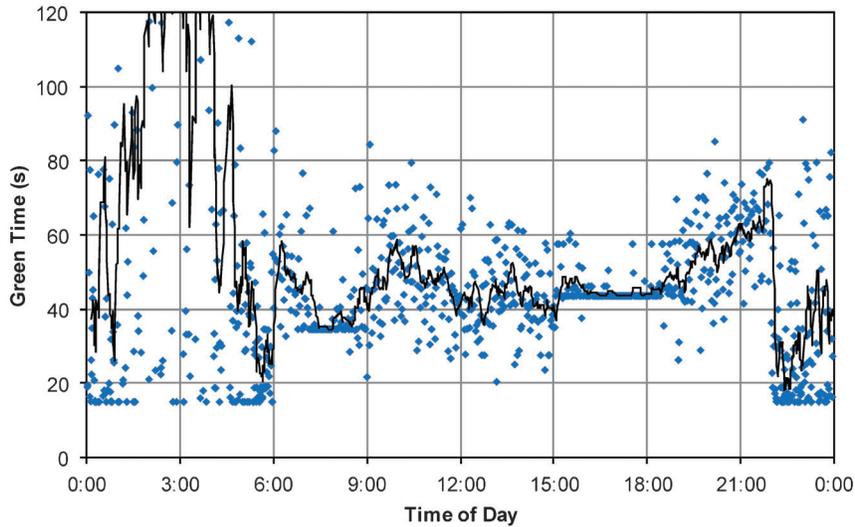


Figure 5.11 Green time (s).

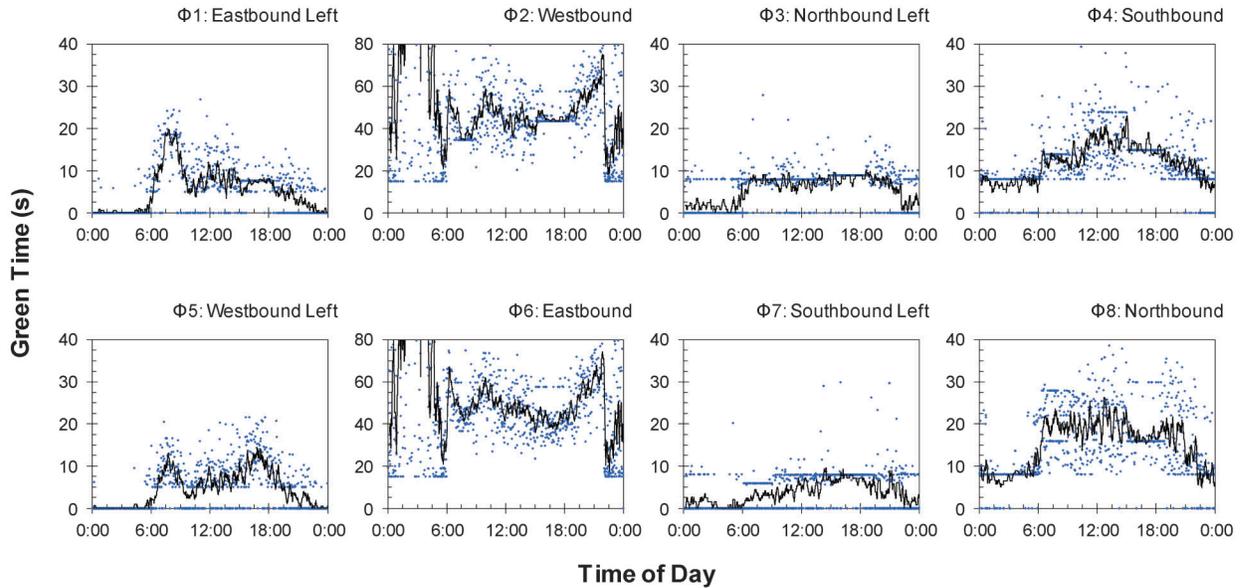


Figure 5.12 Green time (s), eight phases.

the k th vehicle, $T_{A,k}$, represents the time that the controller registered an activation of the advance detector. The estimated arrival time at the stop bar, $T_{a,k}$, is obtained by adding the expected travel time between the detector and the stop bar. In this case, an advance detector is available, and the travel time is five seconds. This aligns the counted vehicles with the state of green upon their arrival at the intersection. The final column shows the service instance j associated with the counted vehicle. Most of these vehicles are counted during the 514th service instance.

These data are compiled into a count per phase interval in Table 5.9. The relevant phase intervals for counting are the two phase 2 end of green times $t_{r,2,j-1}$ (for instance $j - 1$) and $t_{r,2,j}$ (for instance j). The

beginning of green time $t_{g,2,j}$ indexes the count to a given cycle, as illustrated in Figure 3.9. The number of vehicles counted in each interval, $N_{2,j}$, is shown in the final column; this is simply the number of count times satisfying the condition $t_{r,2,j-1} \leq t_{a,k} < t_{r,2,j}$. The equivalent hourly volume $v_{2,j}$ is calculated using Equation 5.4.

As was previously done with the green time metrics, the volumes of the individual service instances are aggregated to cycles. All of the service instances of a given phase occurring during a cycle are summed; usually, this means 0 or 1 instance(s), but 2 or more is sometimes possible. For phase 2, there is exactly one service instance per cycle throughout the day, so $i = j$, as can be seen in Table 5.9. Table 5.10 shows the

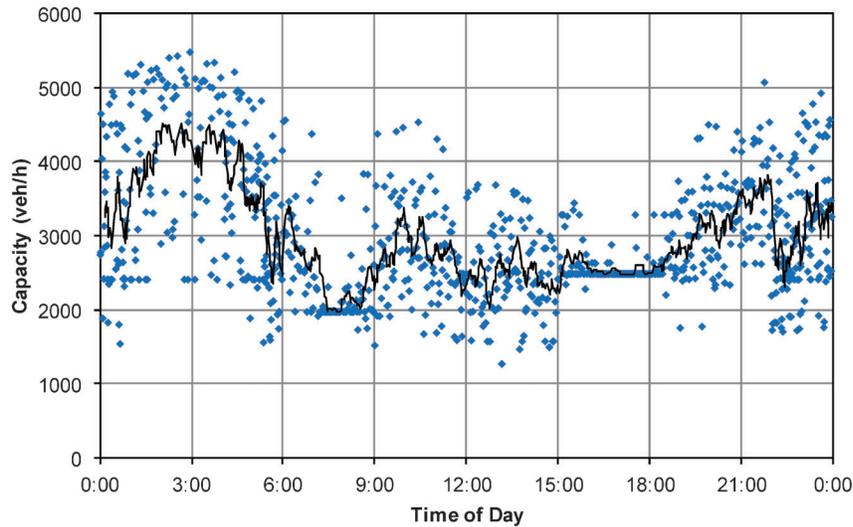


Figure 5.13 Phase 2 capacity (veh/h).

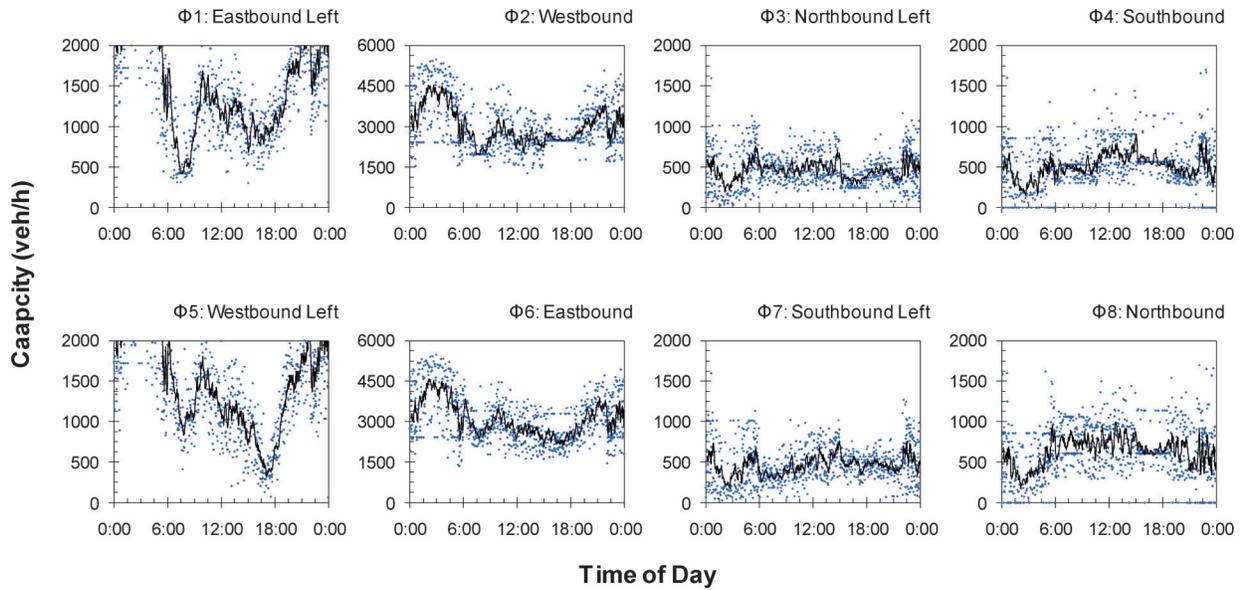


Figure 5.14 Capacity (veh/h), eight phases.

counts compiled for each cycle, along with the capacity (in number of vehicles). The volume-to-capacity ratio $x_{2,j}$ is computed using Equation 5.10.

The next figure series presents example views of the count data and phase utilization using the various metrics calculated in the above tables. As before, an example for phase 2 is provided as well as a plot for all eight phases; the lines in these plots are 10-cycle moving averages. Figure 5.17 and Figure 5.18 show the raw vehicle counts N ; Figure 5.19 and Figure 5.20 show the equivalent hourly volume v ; and Figure 5.21 and Figure 5.22 show the volume-to-capacity ratio x .

The vehicle counts (Figure 5.17 and Figure 5.18) and equivalent hourly volumes (Figure 5.19 and Figure 5.20)

are for the most part very similar. The time-of-day trends are clearly visible in these graphics, especially for the coordinated phases 2 and 6. Less prominent patterns in phases 4 and 8 can also be seen. At some times of day, there are differences between the plots of N and v , which relate to the duration of the counting intervals that affects the scaling in Equation 5.4 (i.e., a shorter C yields a higher v). There are some times of day where a relatively small vehicle count corresponds to a relatively larger equivalent hourly volume, particularly during the early morning and late night shoulder time periods. These plots are also similar in trend to the 15-minute volumes shown earlier in Figure 5.2 and Figure 5.3, as expected. Some individual cycles have volumes

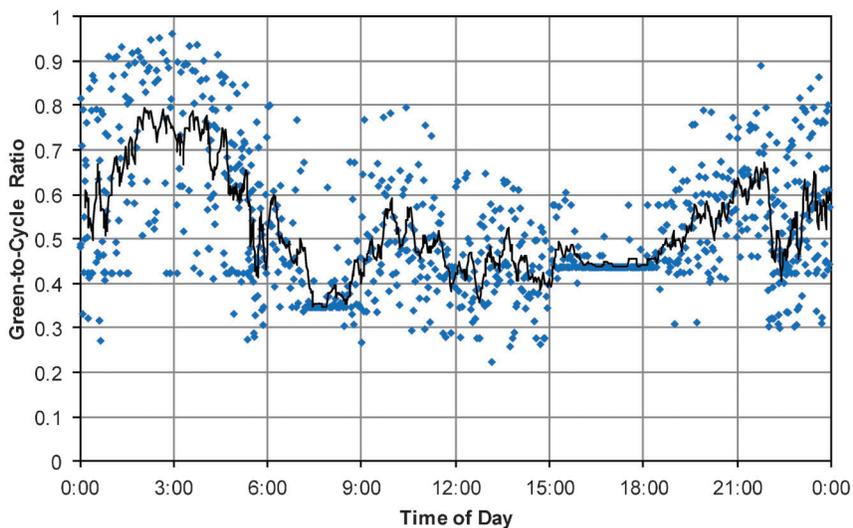


Figure 5.15 g/C ratio.

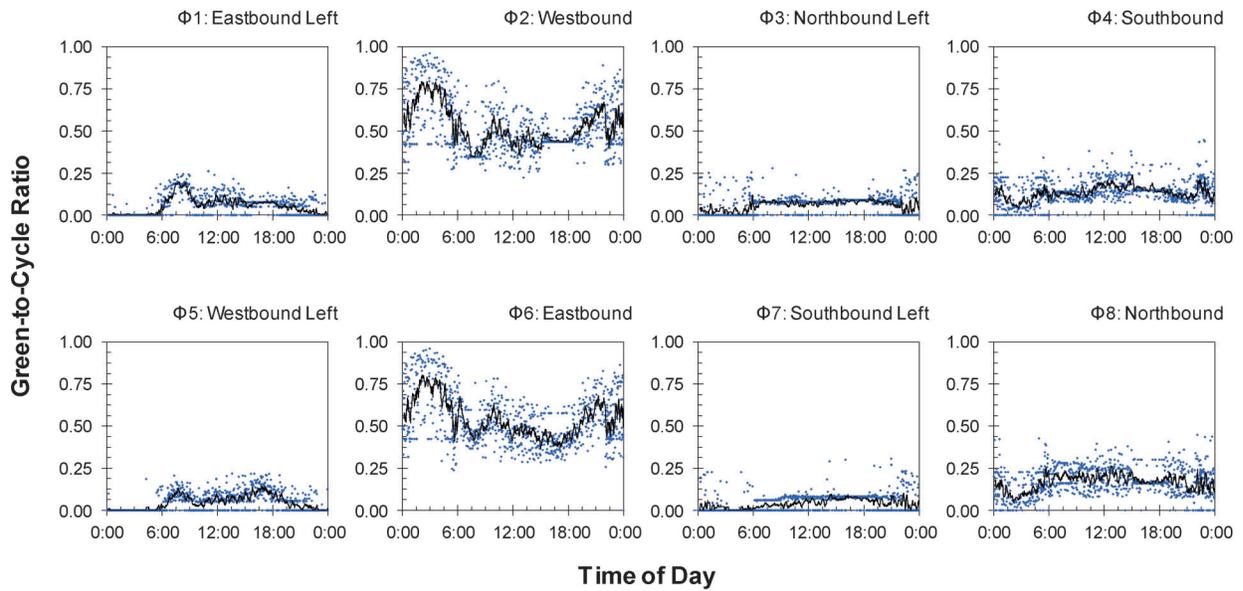


Figure 5.16 g/C ratio, eight phases.

TABLE 5.8
Example vehicle count data (phase 2).

Date	Lane	k	$T_{A,k}$	$T_{a,k}$	j
10/17/2012	A	8513	13:32:24.2	18:32:29.2	513
10/17/2012	B	8514	13:32:46.6	18:32:51.6	514
10/17/2012	A	8515	13:32:58.7	18:33:03.7	514
10/17/2012	B	8516	13:32:59.7	18:33:04.7	514
10/17/2012	A	8517	13:33:01.2	18:33:06.2	514
10/17/2012	A	8518	13:33:03.7	18:33:08.7	514
10/17/2012	B	8519	13:33:06.7	18:33:11.7	514
10/17/2012	B	8520	13:33:09.9	18:33:14.9	514
10/17/2012	A	8521	13:33:12.3	18:33:17.3	514
10/17/2012	A	8522	13:33:12.8	18:33:17.8	514
10/17/2012	B	8523	13:33:15.2	18:33:20.2	514
10/17/2012	A	8524	13:33:17.0	18:33:22.0	514
10/17/2012	B	8525	13:33:17.2	18:33:22.2	514
10/17/2012	A	8526	13:33:20.4	18:33:25.4	514
10/17/2012	A	8527	13:33:27.2	18:33:32.2	514
10/17/2012	B	8528	13:33:31.9	18:33:36.9	514
10/17/2012	B	8529	13:33:33.2	18:33:38.2	514
10/17/2012	C	8530	13:33:40.2	18:33:45.2	514
10/17/2012	A	8531	13:33:43.8	18:33:48.8	514
10/17/2012	A	8532	13:33:48.5	18:33:53.5	514

TABLE 5.9
 Example cycle-by-cycle vehicle count and equivalent hourly volume (phase 2).

Date	$t_{BOC,i}$	$t_{G,2,j}$	i	C_i	φ	j	$N_{2,j}$	$v_{2,j}$
10/17/2012	13:30:17.0	13:31:06.6	512	100.0	2	512	23	828.0
10/17/2012	13:31:57.0	13:32:26.9	513	98.6	2	513	19	730.2
10/17/2012	13:33:36.4	13:34:07.3	514	101.4	2	514	41	1,420.1
10/17/2012	13:35:17.0	13:36:01.2	515	91.6	2	515	23	903.9
10/17/2012	13:36:49.3	13:37:42.9	516	108.4	2	516	26	863.5
10/17/2012	13:38:37.0	13:39:07.8	517	100.0	2	517	22	792.0

TABLE 5.10
 Example data table for volume-to-capacity ratio calculation (phase 2).

Date	$t_{BOC,i}$	$t_{G,2,j}$	i	φ	$N_{2,j}$	$c_{2,i}$	$x_{2,i}$
10/17/2012	13:30:17.0	13:31:06.6	512	2	23	69.7	0.330
10/17/2012	13:31:57.0	13:32:26.9	513	2	19	98.6	0.193
10/17/2012	13:33:36.3	13:34:07.3	514	2	41	100.2	0.409
10/17/2012	13:35:17.0	13:36:01.2	515	2	23	64.9	0.354
10/17/2012	13:36:49.6	13:37:42.9	516	2	26	75.5	0.344
10/17/2012	13:38:37.0	13:39:07.8	517	2	22	99.4	0.221

considerably greater or less than the corresponding 15-minute count, especially during the peaks; this illustrates the stochastic nature of vehicle arrivals.

The volume-to-capacity ratio x joins together the capacity allocation metrics and the served volume metrics to reveal the degree to which the capacity is actually utilized. Although there are few surprises in the trends in x for phases 2 and 6, the other phases reveal times of day when the provided capacity is fully utilized, which cannot be seen in any of the previous plots. For example, phase 1 experiences a strong a.m. peak that is seen as only a modest peak in Figure 5.2, Figure 5.17, or Figure 5.19. If one compares phases 4 and 8 in Figure 5.22, it is clear that phase 4 has a much higher degree of utilization. In contrast, these phases look quite similar in Figure 5.19, except for reaching

peaks at different times of day. Although both phases have peak hour volumes of approximately 500 veh/h, the phase 4 peak occurs when less capacity is available to serve it, so its volume-to-capacity ratio is higher. The volume-to-capacity ratio is more informative than the indicators of capacity allocation or served volumes examined independently.

5.6 Capacity Allocation and Utilization for Protected-Permitted Phases

As mentioned earlier, protected-permitted phases require some additional steps in their calculation processes because the permitted portions of these phases are influenced by another phase at the intersection. To compute permitted phase performance measures, it is

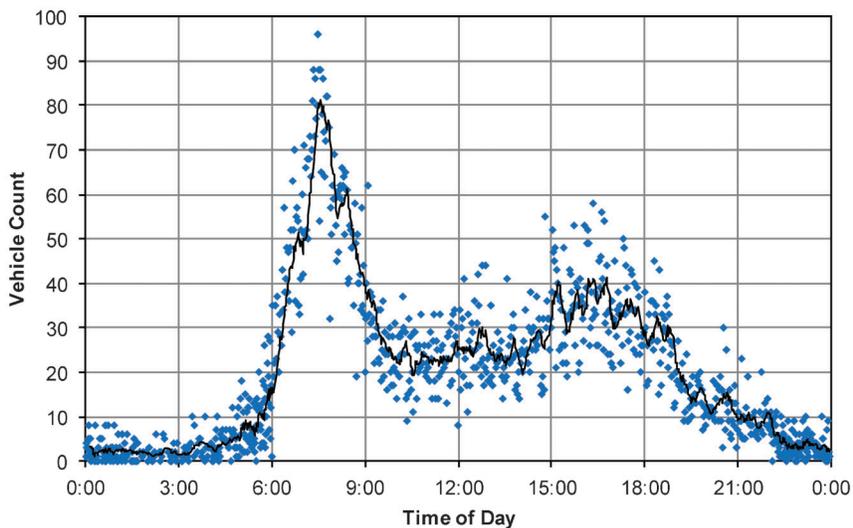


Figure 5.17 Vehicle count.

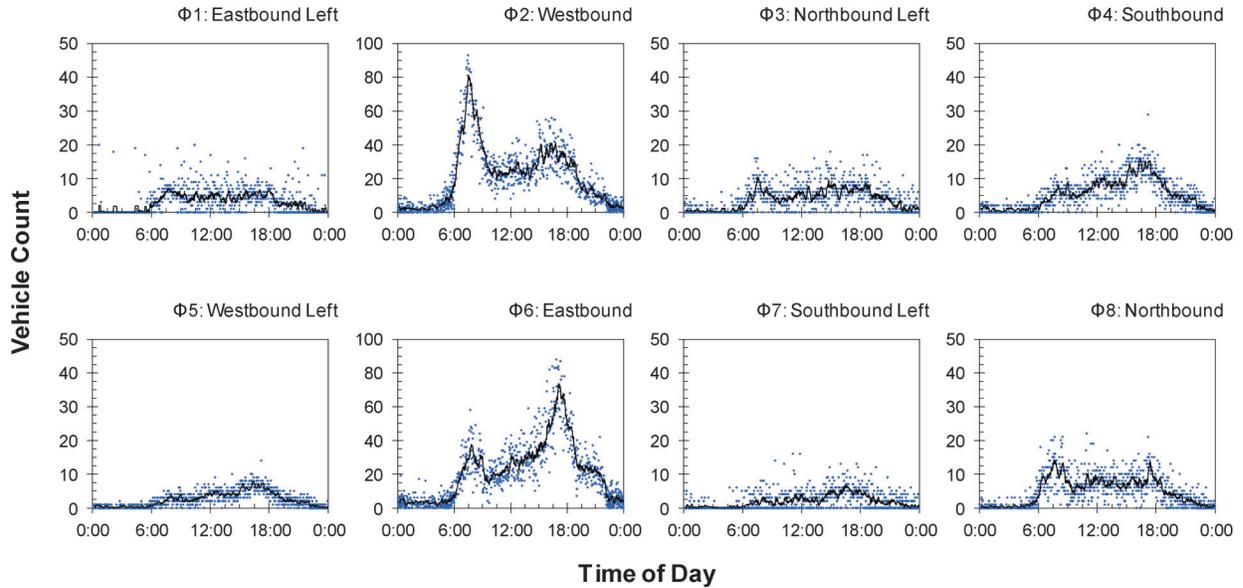


Figure 5.18 Vehicle count, eight phases.

first necessary to calculate the performance measures for the protected phases at the intersection occurring during the same cycle, as these will influence the permitted phase capacity.

The relationship between a protected-permitted left-turn and protected phases was detailed previously in Figure 3.6, Figure 3.7, and Figure 3.8. In summary, the duration of the permitted “green” interval, g_{PERM} , is equal to the minimum of the adjacent through green (g_{AT}) or opposite through movement green (g_{OT}):

$$g_{\text{PERM}} = \min(g_{\text{AT}}, g_{\text{OT}}). \quad (5.12)$$

From this, the permitted capacity c_{PERM} can be calculated:

$$c_{\text{PERM}} = s g_{\text{PERM}} [1 - \min(x_{\text{OT}}, 1)]. \quad (5.13)$$

Here, s is the saturation flow rate of the left-turn movement. The minimum function is needed to filter out cases where $x_{\text{OT}} > 1$ and prevent the calculation of negative capacities. This assumes that the full amount of capacity provided in the permitted phase is reduced by the level of utilization of the opposing through movement.

The capacity of the movement is given by

$$c = c_{\text{PROT}} + c_{\text{PERM}} + c_{\text{SNEAK}}. \quad (5.14)$$

In this equation, c_{PROT} is the capacity of the protected phase, which is calculated by Equation 5.5, and c_{SNEAK} is the capacity of what is sometimes called the “sneakers” who execute the left-turn movement at the end of the permitted interval. When all of the

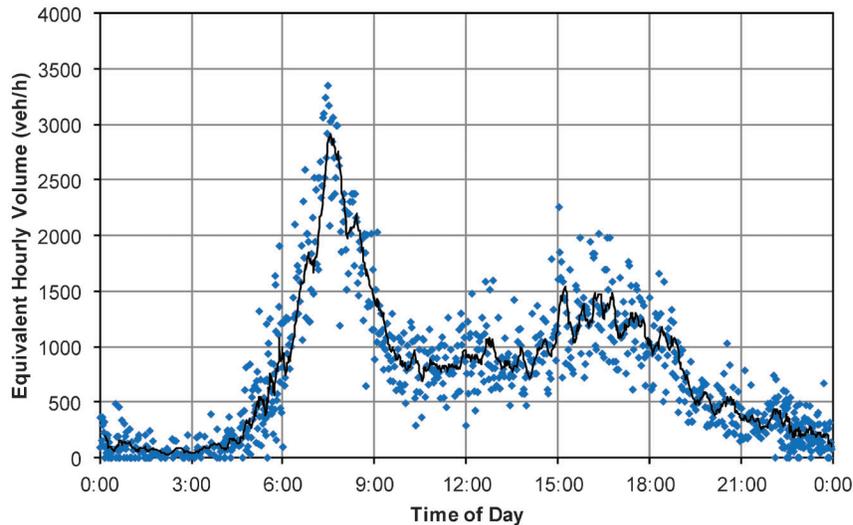


Figure 5.19 Equivalent hourly volume.

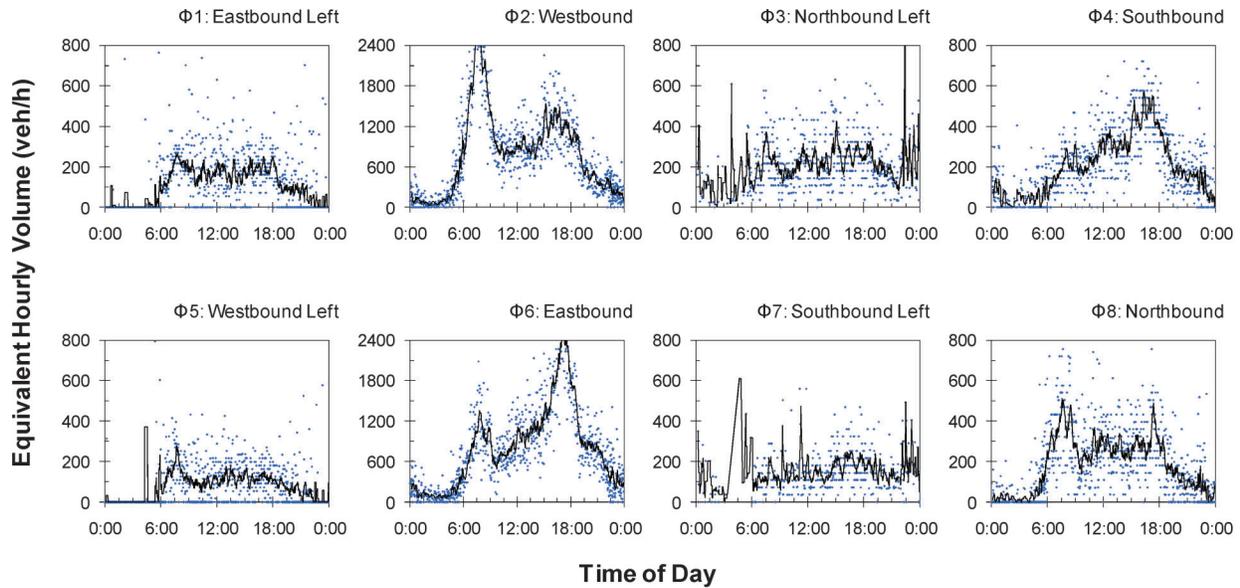


Figure 5.20 Equivalent hourly volume.

capacities are expressed in units of vehicles, a value of 1 or 2 vehicles is reasonable for c_{SNEAK} . This implies that there is always a certain amount of capacity provided for the left-turn movements, even if the protected phase is omitted and the opposing through movement is at capacity.

An example real-world situation is presented in Figure 5.23. This shows the geographic situation of the northbound left turn, controlled by phase 3, at US 36 and Post Road. The opposing through movement, controlled by phase 4, crosses the path of the northbound left. Therefore, the utilization of phase 4 controls the permitted capacity of phase 3. Because this phase operates a five-section head (rather than a flashing yellow arrow), the permitted interval occurs during the green indication of the adjacent through movement, phase 8.

Figure 5.24 shows the amounts of protected and permitted capacity associated with phase 3 throughout the day on October 17, 2012. The capacities have been converted to vehicles per hour using Equation 5.6. A value of $c_{\text{SNEAK}} = 2$ means that every cycle has a capacity of no less than 2 vehicles. The permitted capacity is constrained during the peak periods, as might be expected. These are the time periods when the served volumes on phase 8 are relatively high.

Table 5.11 provides an example of the calculation of capacity and volume-to-capacity ratio for phase 3. Table 5.11a, Table 5.11b, and Table 5.11c present the relevant information for phases 3, 8, and 4, respectively. The duration of protected green of phase 3 (g_3), and the durations of phases 4 and 8 (g_4 and g_8) are needed. The volume-to-capacity ratio of phase 4 (x_4) is also needed. Table 5.11d shows the details of the calculation for six

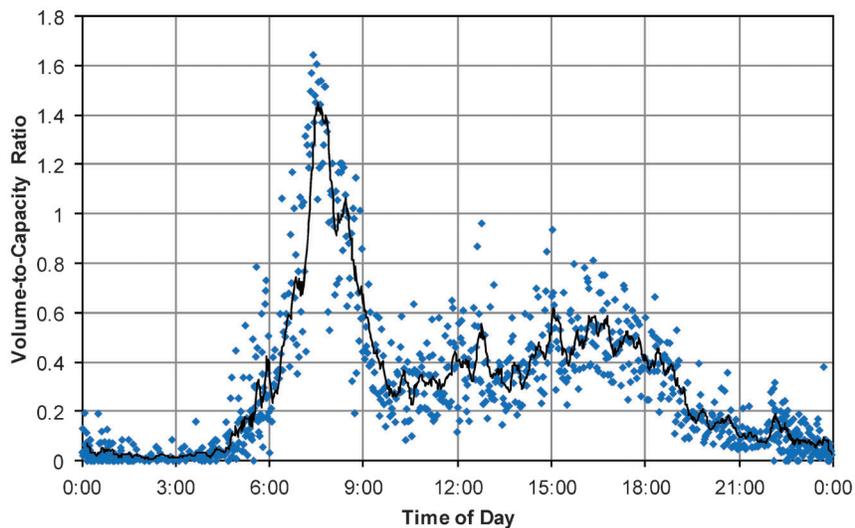


Figure 5.21 Volume-to-capacity ratio.

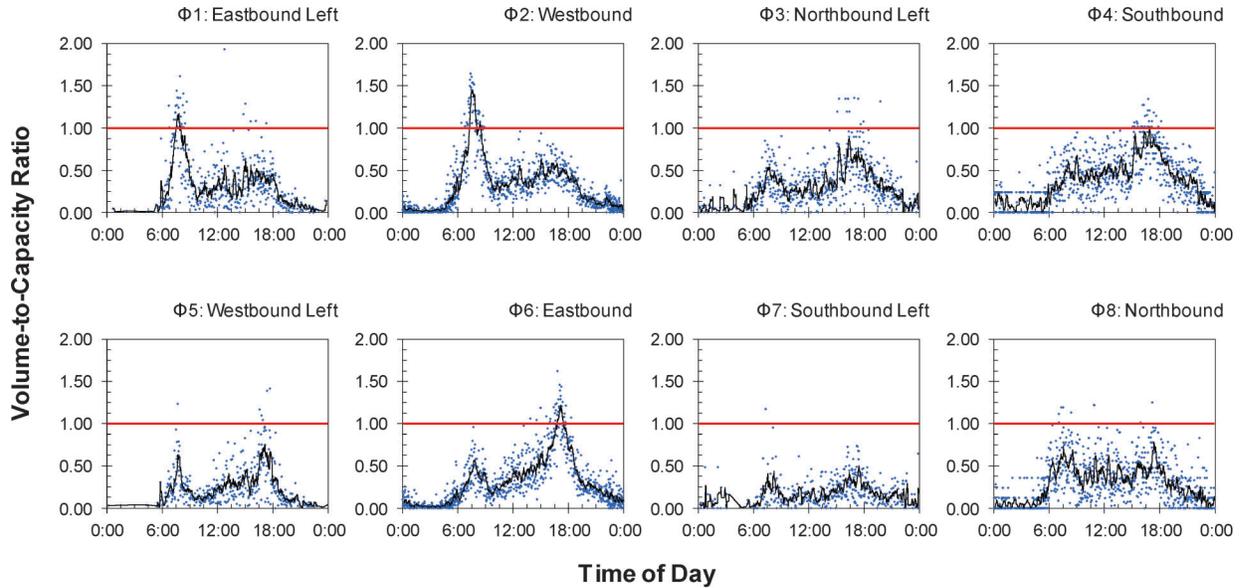


Figure 5.22 Volume-to-capacity ratio, eight phases.

cycles during the afternoon of October 17, 2012. The data shows the sensitivity of c_3 , and therefore x_3 , on the utilization of the opposing through movement x_4 . During cycle $i = 475$, eight vehicles were served under a capacity of 10.5 vehicles, yielding $x_3 = 0.761$. During cycle $i = 480$, ten vehicles were served within a similar amount of permitted green time (g_{PERM}), but the

opposing through movement utilization was lower, so the total available capacity was 19.6 and $x_3 = 0.509$.

5.7 Phase Termination

The reason for each phase terminations, when tabulated, can be used to effectively characterize capacity utilization (4,45). There are four possible reasons for phase termination in a conventional controller. An actuated phase can be *omitted* when there is no call for the phase during a cycle; it can *gap-out* when the minimum green time has been served, and the vehicle extension timer have expired (indicating that the demand has been served); it can *max-out* when the controller is non-coordinated and the phase time has been extended to the maximum limit under conflicting demand; and it can *force-off* when the controller is running under coordinated operation and the split timer expires. Figure 5.25 gives an example of phase termination event occurrences at an eight-phase intersection over a 24-hour period.

Max-outs and force-offs indicate that a phase is exceeding capacity, while gap-outs and omits indicate that there is capacity to spare. Within each cycle, if it is found that certain phases are maxing out while others are gapping out or are omitted, it is likely that capacity could be exchanged between phases. In a typical eight-phase ring diagram (Figure 5.26), there are five potential locations where split times can be exchanged. All members of the four sets of “companion” phases $\{1,2\}$, $\{3,4\}$, $\{5,6\}$, and $\{7,8\}$ are physically conflicting, contained in one ring, and belong to the same phase group. The controller must divide time between the two companion phases. The four possible split exchanges $\{\delta_1, \delta_2, \delta_3, \delta_4\}$ existing within the four pairs represent potential locations for adjustment opportunities. A fifth possible exchange Δ exists between the two phase groups, where split time would be collectively transferred between $\{1, 2, 5, 6\}$ and $\{3, 4, 7, 8\}$. In general,

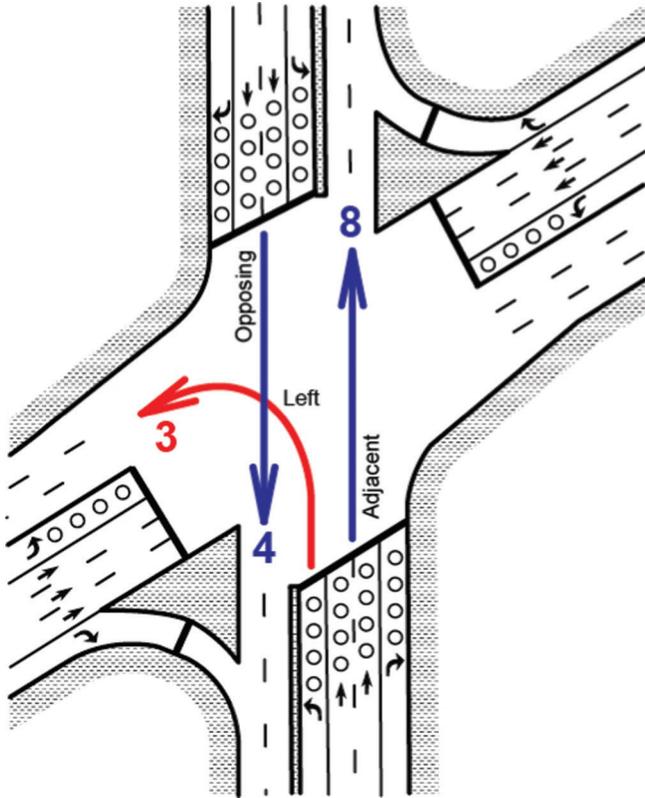


Figure 5.23 Through phases related to capacity calculation for phase 3.

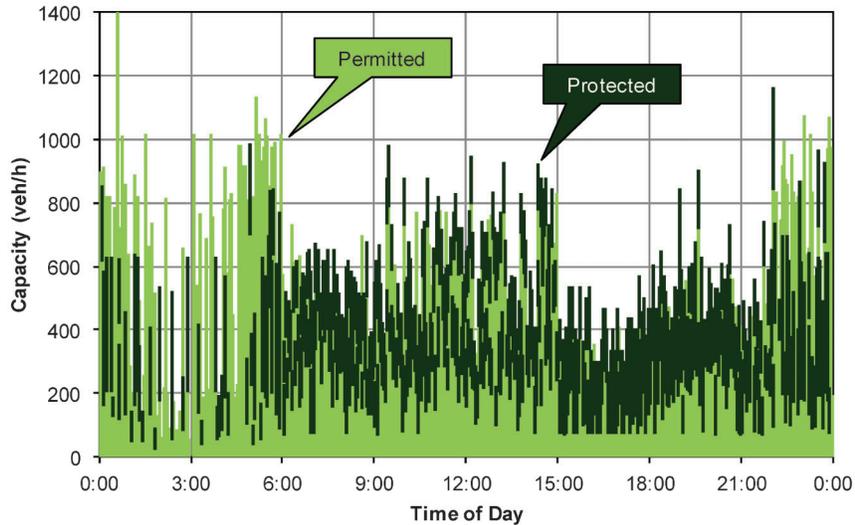


Figure 5.24 Phase 3 capacity (veh/h): Protected and permitted components.

any set of phases can be analyzed using this method, if the phases cannot be run simultaneously and must share a given amount of capacity in a cycle.

Within any pair of companion phases, there is a leading phase and a lagging phase. In Figure 5.26, for example, phase 1 is leading and phase 2 is lagging. Phase actuation tends to favor the lagging phase, because when the leading phase gaps out, the lagging phase inherits the time (particularly with fixed force-off). This is represented in Figure 5.26 with the dashed lines on the right side of the split exchanges. While some caution is advisable to avoid excessively loading the leading phases with split time, in general we are interested in identifying opportunities to move split time from lagging phases to leading phases rather than vice versa, because the non-actuation and subsequent gap-out of one phase during low demand will transfer the time to the next phase in the ring.

Figure 5.27 shows phase terminations for all eight phases of an intersection over a three-hour period. The hollow circles indicate a force-off event that was accompanied by a gap-out event of the companion phase, within the same cycle. The solid circles indicate the same force-off and gap-out condition that has appeared for three consecutive cycles. Generally, we would expect random occurrences of excess demand to be corrected in one or two cycles, and filtering by three consecutive cycles can mitigate the effects of stochastic demand.

5.8 Green Occupancy Ratio

Another measure of phase utilization is the amount of stop bar detector occupancy taking place during green, or the *green occupancy ratio* (GOR). This is defined by

$$\text{GOR} = \frac{T_{\text{ON}}}{T_{\text{ON}} + T_{\text{OFF}}}, \quad (5.15)$$

where T_{ON} and T_{OFF} are the total durations of detector ON and OFF times during the green interval. This performance measure supports adaptive split

control in ACS-Lite (46) and can be applied wherever stop bar detection is available. There is, however, some sensitivity of GOR to the length of the detection zone (47). The longer the detection zone, the greater the tendency for larger GOR values to be calculated. Many intersections do not feature stop bar detection on all phases. For example, in the state of Indiana, the mainline coordinated movements usually feature advance detectors, but stop bar detectors are not available for those movements.

Figure 5.28 shows a plot of GOR for phase 3 at US 36 and Post Road on October 17, 2012. The data shows a rather bimodal characteristic. Throughout most of the day, GOR ranges between approximately 0.1 and 0.4, but during the peak periods it sporadically increases to above 0.5. This is a typical profile of occupancy, which tends to reach a maximum value relatively quickly as volume begins to increase. For comparison, an enlarged plot of the volume-to-capacity ratio is provided in Figure 5.29. Note that the peaks have greater width and a more gradual rise. There is still considerable variation in the volume-to-capacity ratio from one cycle to the next, as a result of stochastic variation in volumes and green times. However, varies less than GOR. Around 17:00–18:00, the central tendency of the volume-to-capacity ratio (shown by the 10-point moving average) is clearly in the range of about 0.6 in Figure 5.29. However, in Figure 5.28, the moving average line for the same time period itself ranges between 0.1 and 0.5, and the point cloud spans the entire range of possible values. GOR has greater value as a threshold-based indicator of high utilization (e.g., $\text{GOR} > 0.8$) than as a detailed measure of the degree of utilization. Nevertheless, GOR can be used under many different presence detector configurations, whereas the volume-to-capacity ratio requires count detection.

5.9 Green Occupancy Ratio and Red Occupancy Ratio

The GOR has been shown to have a tendency to reach a high value earlier than the volume-to-capacity

TABLE 5.11
Protected and permitted capacity calculation. Volumes and capacities are shown as number of vehicles.

(a) Phase 3 (left-turn) data.													
Date	$t_{BOC,i}$	i	ϕ	$t_{G,3j}$	g_3								
10/17/2012	12:28:37.0	475	3	12:28:37.0	7.9								
10/17/2012	12:30:17.0	476	3	12:30:17.0	7.9								
10/17/2012	12:31:57.0	477	3	12:31:57.0	7.9								
10/17/2012	12:33:37.0	478	3	12:33:37.0	7.9								
10/17/2012	12:35:17.0	479	3	12:35:17.0	7.9								
10/17/2012	12:36:57.0	480	3	12:36:57.0	7.9								
(b) Phase 8 (adjacent through) data.													
Date	$t_{BOC,i}$	i	ϕ	$t_{G,8j}$	g_8								
10/17/2012	12:28:37.0	475	8	12:28:37.0	35.4								
10/17/2012	12:30:17.0	476	8	12:30:31.0	23.9								
10/17/2012	12:31:57.0	477	8	12:32:11.0	17.7								
10/17/2012	12:33:37.0	478	8	12:33:51.0	22.4								
10/17/2012	12:35:17.0	479	8	12:35:17.0	30.2								
10/17/2012	12:36:57.0	480	8	12:37:11.0	23.9								
(c) Phase 4 (opposing through) data.													
Date	$t_{BOC,i}$	i	ϕ	$t_{G,4j}$	g_4	x_4							
10/17/2012	12:28:37.0	475	4	12:28:51.0	21.4	0.797							
10/17/2012	12:30:17.0	476	4	12:30:31.0	23.9	0.396							
10/17/2012	12:31:57.0	477	4	12:32:11.0	17.7	0.482							
10/17/2012	12:33:37.0	478	4	12:33:51.0	22.4	0.423							
10/17/2012	12:35:17.0	479	4	12:35:31.0	16.2	0.760							
10/17/2012	12:36:57.0	480	4	12:37:11.0	23.9	0.436							
(d) Final computation of phase 3 volume-to-capacity ratio.													
i	g_{OT} (g_4)	g_{AT} (g_8)	g_{PERM}	g_{PROT} (g_3)	c_{PROT}	x_{OT} (x_4)	c_{PERM}	c_{SNEAK}	N_{PERM}	N_{PROT}	N_3	c_3	x_3
475	21.4	35.4	21.4	7.9	4.2	0.797	6.3	2.0	3	5	8	10.5	0.761
476	23.9	23.9	23.9	7.9	4.2	0.396	16.4	2.0	2	2	4	20.6	0.194
477	17.7	17.7	17.7	7.9	4.2	0.482	11.2	2.0	0	5	5	15.3	0.326
478	22.4	22.4	22.4	7.9	4.2	0.423	14.9	2.0	2	7	9	19.1	0.471
479	16.2	30.2	16.2	7.9	4.2	0.760	5.9	2.0	2	2	4	10.1	0.398
480	23.9	23.9	23.9	7.9	4.2	0.436	15.5	2.0	5	5	10	19.6	0.509

ratio, thereby making it less effective as a continuous variable for accurately measuring the degree of utilization, and more effective as a threshold-based indicator of high utilization. High utilization alone, however, is not necessarily an indicator of poor phase operation. Actually, for actuated phases, it is desirable to terminate the phase shortly after the demand has been served; it is not efficient to dwell in green for long intervals when no vehicles are being served and yet there is demand for service on other phases. What is undesirable, however, is for a split failure to occur. A split failure occurs when a phase is unable to serve all of its demand within one signal cycle. Vehicles affected by a split failure will accrue a much higher amount of delay than those that only had to wait through one cycle.

One way of measuring whether there is leftover demand at the end of a cycle is to take a look at detector occupancy at the beginning of red—a *red*

occupancy ratio (ROR). The ROR during the first five seconds of red, ROR_5 , is defined as (48)

$$ROR_5 = \frac{T_{ON}}{5}, \quad (5.16)$$

where T_{ON} is the detector occupancy during the first five seconds of red. In this case, the beginning of red is defined as the actual beginning of the red indication. The yellow interval is not used for either GOR or ROR_5 owing to the uncertainty of the utilization of yellow for vehicle movement.

Just as GOR by itself does not necessarily indicate a split failure, ROR_5 merely shows that vehicles were present after the end of green: They might have simply arrived at that time, or they may not have been part of the queue of vehicles waiting for service at the beginning of green. Combining GOR and ROR_5 into a composite performance measure allows for a more powerful evaluation of split performance than either

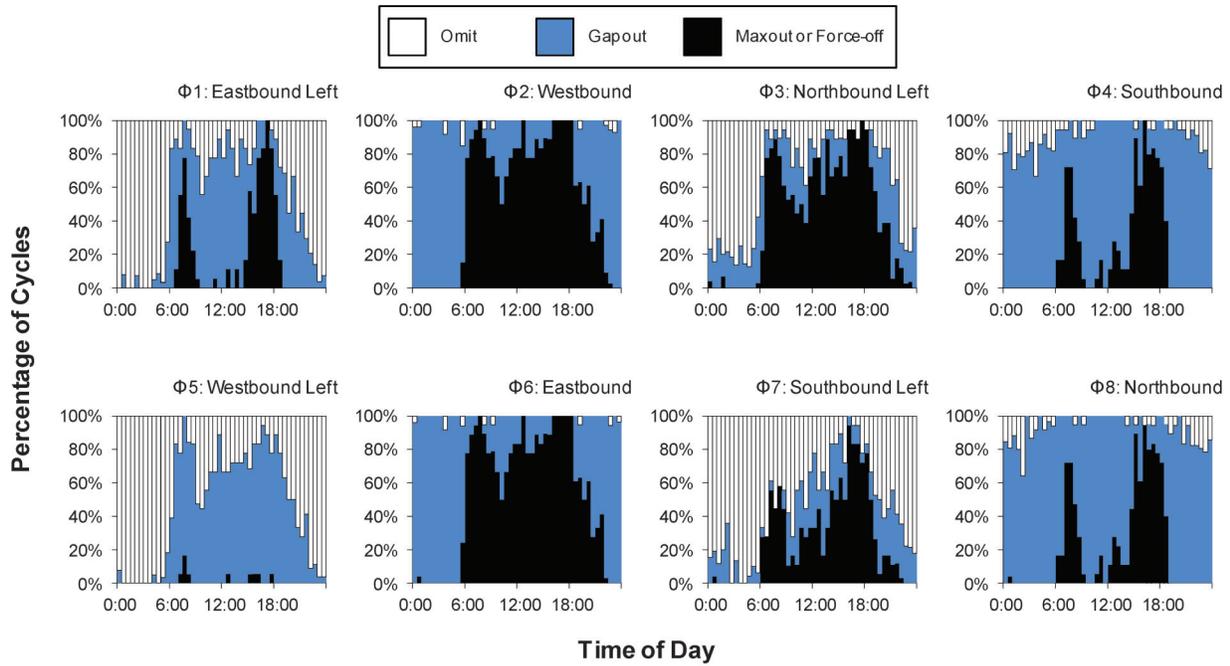


Figure 5.25 Distributions of reasons for phase termination by half-hour bin.

performance measure separately. When both GOR and ROR_5 are high, it indicates not only that high utilization occurred, but that there was also leftover demand after the end of green.

A recent pilot study (48) on this performance measure was carried out for data from the intersection of US 30 and 116th Street in Carmel, Indiana. Figure 5.30 shows a map of this intersection and a view of its phase assignment and locations of detectors.

Figure 5.31 shows an example plot of ROR_5 versus GOR for eastbound phase 4 at the study intersection. Each point in this plot represents the result for one signal cycle during the 09:00–15:00 time period. The circles represent gap-outs, and the diamonds represent force-offs. The quadrant of the plot in which each symbol lies allows characterization of the operation of the phase.

- Symbols in the upper right quadrant of the plot represent likely split failures, especially the force-offs indicated by the diamonds, where $ROR_5 \geq 80$ and $GOR \geq 80$. The gap-outs (circles) indicate phases that likely terminated because of low demand (it is also possible that very aggressive gap times might cause a gap-out during saturation flow conditions).
- Symbols in the lower right quadrant indicate phases represent conditions that are near saturation. However, the high GOR and zero ROR_5 suggest efficient phase operation rather than a split failure, since there are clearly no vehicles remaining at the beginning of red. A low value of ROR_5 may be attributable to a late arrival or vehicle crossing the stop bar as the indication turns red. This is especially clear when the phase also gaps out.
- The upper and lower left quadrants represent under-saturated conditions, as indicated by the low values of GOR.

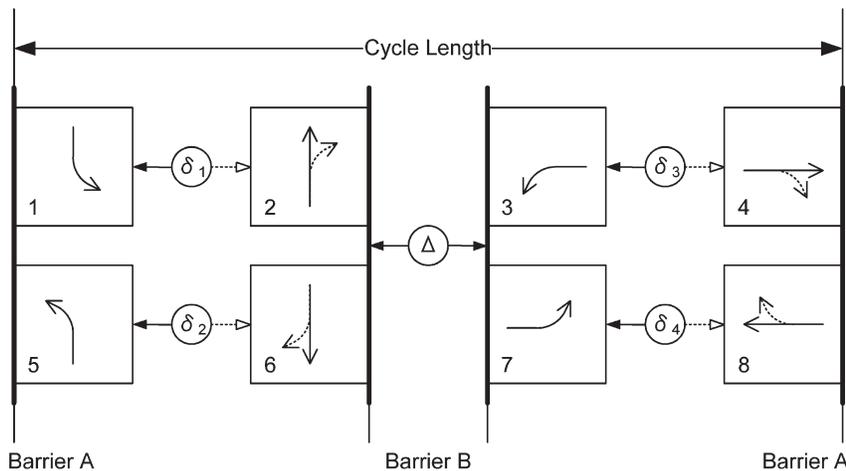


Figure 5.26 Definitions of barrier crossing times (45).

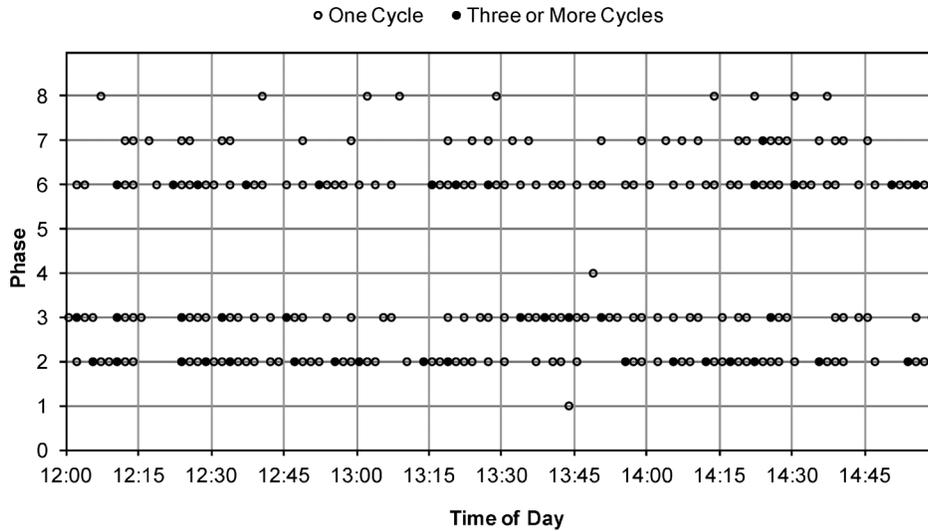


Figure 5.27 Phase termination diagram showing time redistribution opportunities (45).

Figure 5.32 shows the same plots for all eight phases at the intersection (Figure 5.32a through Figure 5.32h). These illustrate how the ROR_5 versus GOR scatterplot can effectively illustrate overall intersection operations. There are clear opportunities for improvement on phases 1, 3, 4, and 8, which have numerous cycles in the upper right quadrant of their plots. On the other hand, phases 5 and 7 are clearly undersaturated.

Phases 2 and 6 do not have stop bar detectors, but instead feature setback detection. Therefore, the ROR_5 and GOR information for these phases is less valid. ROR_5 derived from setback detectors does not represent the occupancy of vehicles stopped at the intersection. This explains why these plots appear quite different from the other phases. Instead, the volume-to-capacity ratio would be used to characterize these phases, as shown in Figure 5.32i and Figure 5.32j.

5.10 Degree of Intersection Saturation

The intersection degree of saturation, X_C , is an overall measure of the utilization of capacity across an intersection. The general formula for X_C is (40,49)

$$X_C = \left(\frac{C}{C-L} \right) \sum_{c\phi} y_{c\phi}, \quad (5.17)$$

where C is the cycle length, L is the total lost time (clearance time), and the sum over “critical” phases $c\phi$ is the sum of flow ratios y in what is called the *critical path* through the intersection. A flow ratio is the proportion of the volume to the saturation flow rate:

$$y = \frac{v}{s}. \quad (5.18)$$

The critical phase concept arises from the fact that in multiple-ring configurations, alternative phase groups can be served concurrently, and must wait at barriers before the controller can proceed with the next phase group. This means that one phase group must wait for the other to terminate. The critical phase group is the one with the greater volume.

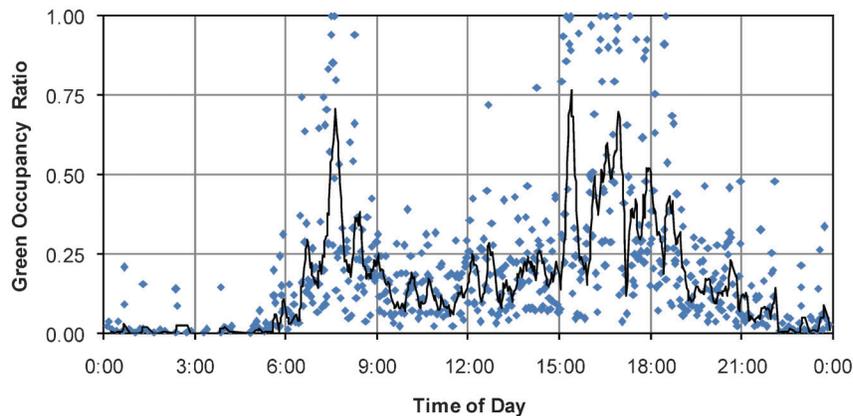


Figure 5.28 Phase 3 green occupancy ratio.

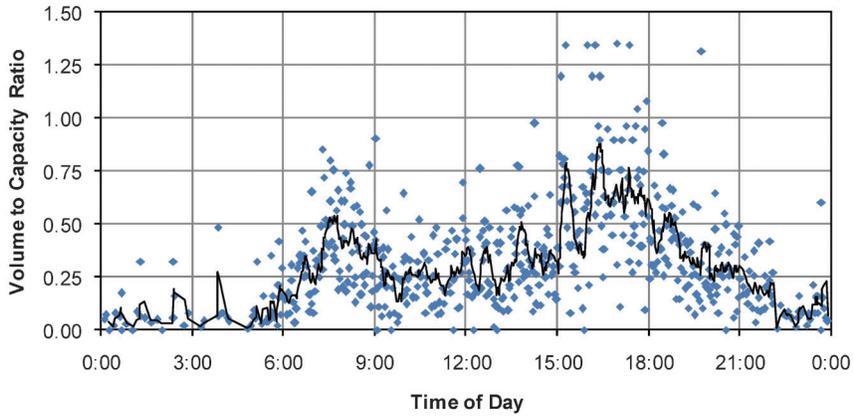


Figure 5.29 Phase 3 volume-to-capacity ratio.

For a standard dual-ring, eight-phase controller, Equation 5.17 has the following practical equivalent (50):

$$X_C = \left(\frac{C}{C-L} \right) \{ \max[(y_1 + y_2), (y_5 + y_6)] + \max[(y_3 + y_4), (y_7 + y_8)] \}. \quad (5.19)$$

The meaning behind this equation is explained in Figure 5.33, which shows the four possible critical paths through the eight-phase, dual-ring control scheme. In the first block of phases {1, 2, 5, 6}, the critical path will pass through whichever group of complementary phases has the greater volume, {1, 2} or {5, 6}. A similar comparison is made between {3, 4} and {7, 8}.

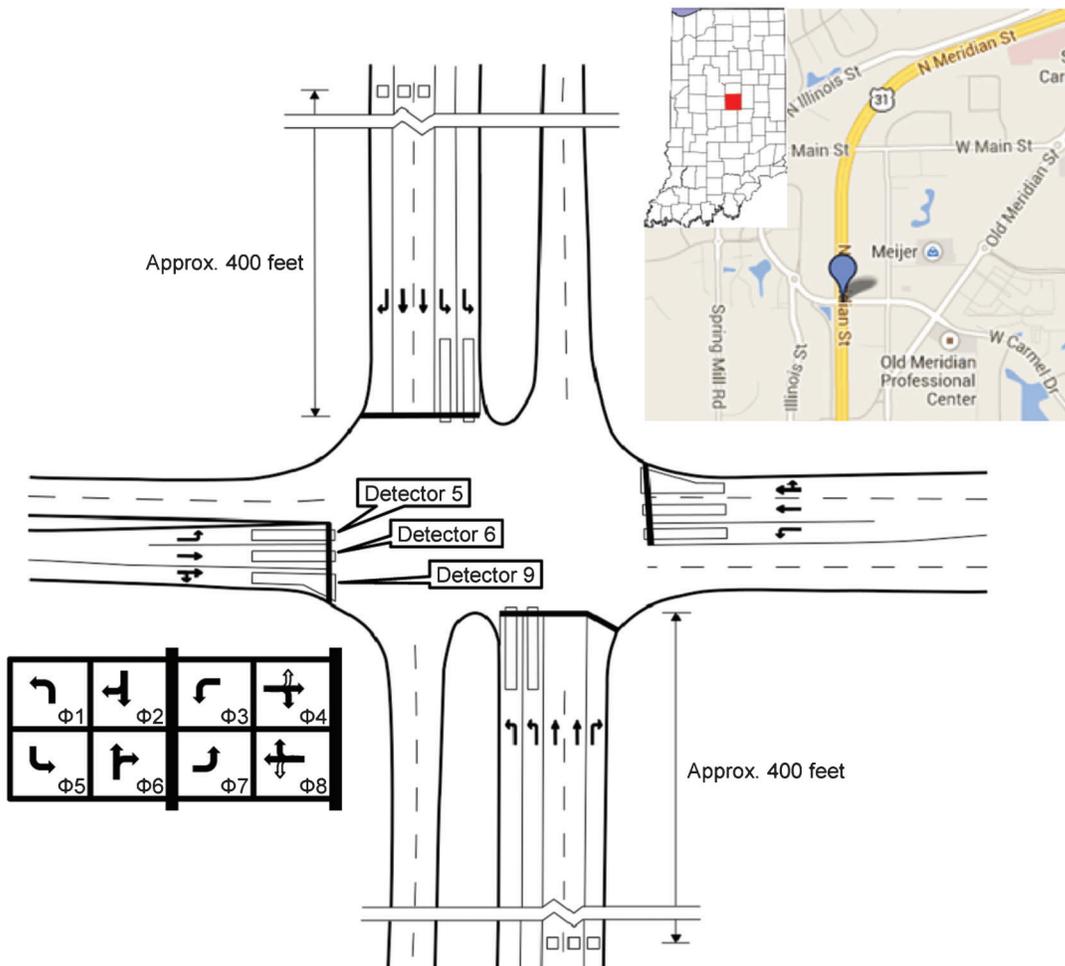


Figure 5.30 The intersection of US 31 and 126th Street in Carmel, Indiana (48).

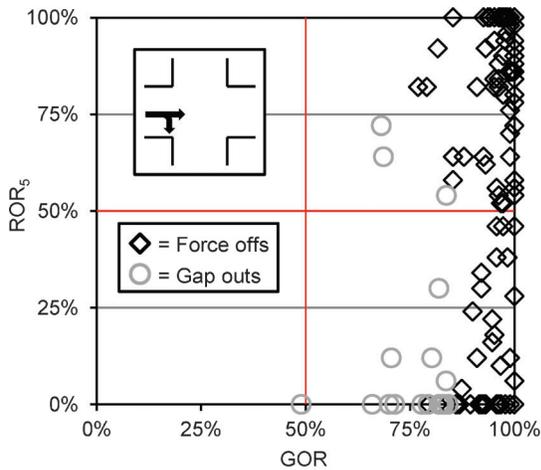


Figure 5.31 ROR₅ versus GOR plot for phase 4 at US 31 and 126th Street (48). The data are tabulated for 09:00–15:00 on June 26, 2013.

The computation of X_C requires that all of the individual volumes for the eight phases have been previously compiled and aggregated by cycle. The components of X_C belonging to the two phase groups {1, 2, 5, 6} and {3, 4, 7, 8} are then computed separately and then summed to obtain the total X_C . Example calculations are shown in three tables. Table 5.12 shows the calculation of the degree of utilization for phase group 1, Table 5.13 shows that for phase group 2, and Table 5.14 shows the final calculation steps. In Table 5.12c, it can be seen that phases {5, 6} have the greater volume in four of the six example cycles, and Table 5.13c reveals that {3, 4} usually have a higher volume than {7, 8}. The impact of actuation can also be seen; when phases are omitted,

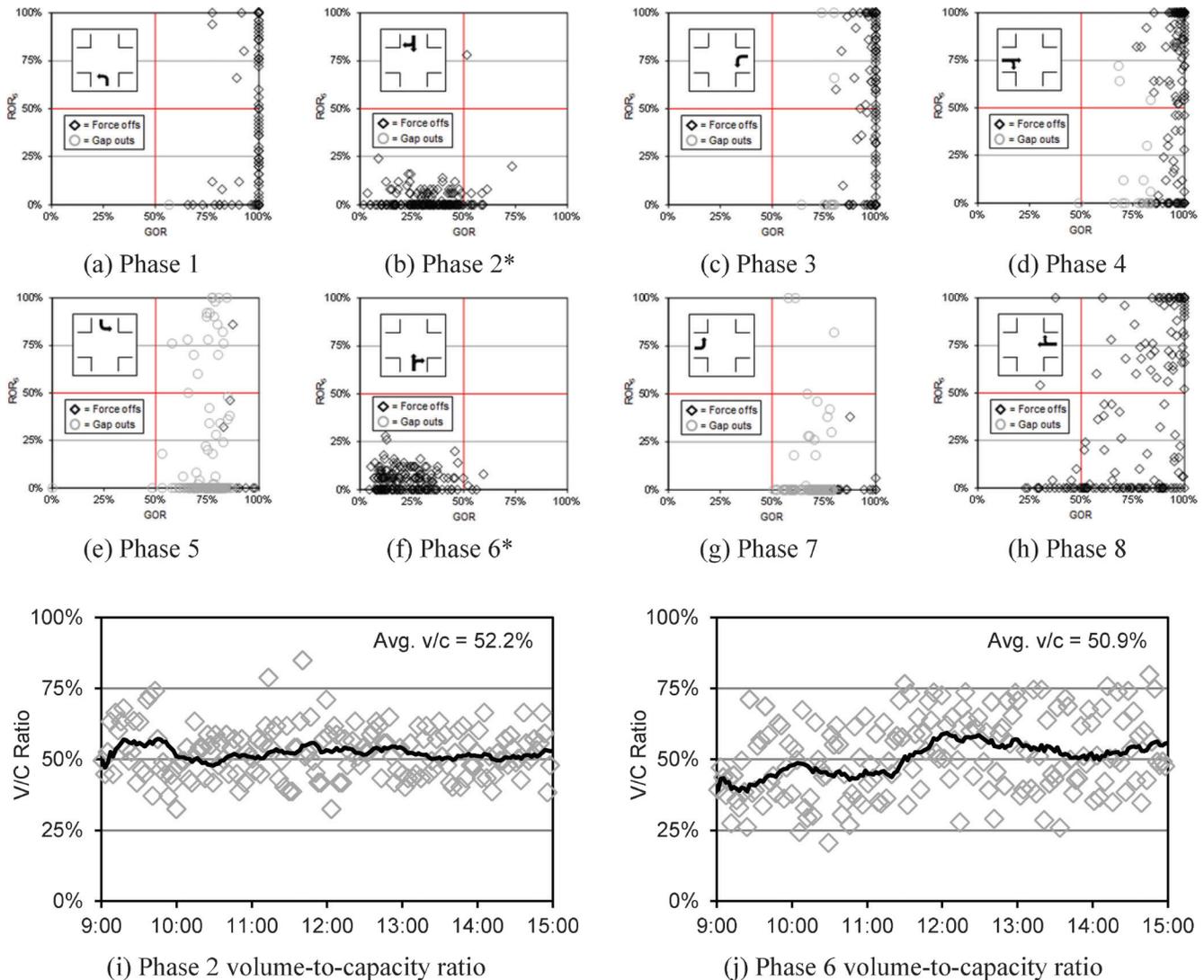


Figure 5.32 ROR₅ versus GOR plots for eight phases, and volume-to-capacity ratio for phases 2 and 6 (48). (*NOTE: Phase 2 and 6 based on advance detector occupancy and not representative of phase utilization.) The data are tabulated for 09:00–15:00 on June 26, 2013.

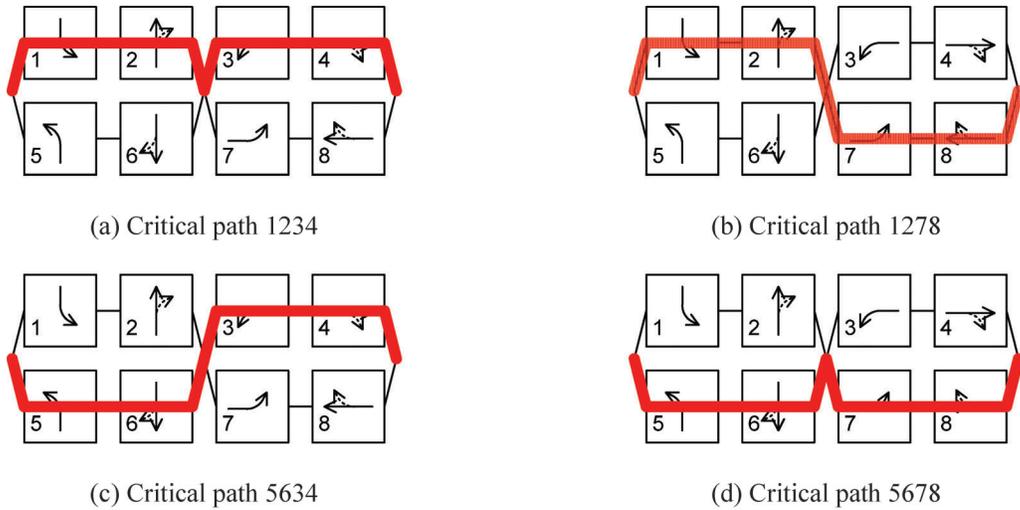


Figure 5.33 Explanation of critical path in eight-phase, dual-ring signal operation.

TABLE 5.12
 X_C calculation, phase group 1.

(a) Ring 1, group 1 (phases 1 and 2).

Date	$t_{BOC,i}$	Phase 1			Phase 2			Lost Time			$(v/s)_{1,1}$
		v_1	s_1	v_1/s_1	v_2	s_2	v_2/s_2	l_1	l_2	$l_{(1+2)}$	
10/17/2012	13:30:17.0	216	1900	0.114	828	5700	0.145	6.4	6.4	12.8	0.259
10/17/2012	13:31:57.0	0	1900	0.000	730	5700	0.128	0.0	6.4	6.4	0.128
10/17/2012	13:33:36.3	0	1900	0.000	1420	5700	0.249	0.0	6.4	6.4	0.249
10/17/2012	13:35:17.0	668	1900	0.352	904	5700	0.159	6.4	6.4	12.8	0.510
10/17/2012	13:36:49.6	199	1900	0.105	863	5700	0.151	6.4	6.4	12.8	0.256
10/17/2012	13:38:37.0	0	1900	0.000	792	5700	0.139	0.0	6.4	6.4	0.139

(b) Ring 2, group 1 (phases 5 and 6).

Date	$t_{BOC,i}$	Phase 1			Phase 2			Lost Time			$(v/s)_{2,1}$
		v_5	s_5	v_5/s_5	v_6	s_6	v_6/s_6	l_5	l_6	$l_{(5+6)}$	
10/17/2012	13:30:17.0	288	1900	0.152	1080	5700	0.189	6.4	6.4	12.8	0.341
10/17/2012	13:31:57.0	0	1900	0.000	986	5700	0.173	6.4	6.4	12.8	0.173
10/17/2012	13:33:36.3	391	1900	0.206	1030	5700	0.181	6.4	6.4	12.8	0.386
10/17/2012	13:35:17.0	432	1900	0.228	1218	5700	0.214	6.4	6.4	12.8	0.441
10/17/2012	13:36:49.6	166	1900	0.087	863	5700	0.151	6.4	6.4	12.8	0.239
10/17/2012	13:38:37.0	0	1900	0.000	900	5700	0.158	6.4	6.4	12.8	0.158

(c) Group 1 critical path determination.

Date	$t_{BOC,i}$	i	Compared Values		Lost Times		Selected Values		
			$(v/s)_{1,1}$	$(v/s)_{2,1}$	$l_{(1+2)}$	$l_{(5+6)}$	Critical Path	Lost Time	$(v/s)_{G1}$
10/17/2012	13:30:17.0	512	0.259	0.341	12.8	12.2	56	12.8	0.341
10/17/2012	13:31:57.0	513	0.128	0.173	6.4	12.2	56	6.4	0.173
10/17/2012	13:33:36.3	514	0.249	0.386	12.8	12.2	56	12.8	0.386
10/17/2012	13:35:17.0	515	0.510	0.441	12.8	12.2	12	12.8	0.510
10/17/2012	13:36:49.6	516	0.256	0.239	12.8	12.2	12	12.8	0.256
10/17/2012	13:38:37.0	517	0.139	0.158	6.4	12.2	56	6.4	0.158

TABLE 5.13
 X_C calculation, phase group 2.

(a) Ring 1, group 2 (phases 3 and 4).											
Date	$t_{BOC,i}$	Phase 1			Phase 2			Lost Time			$(v/s)_{1,2}$
		v_3	s_3	v_3/s_3	v_4	s_4	v_4/s_4	l_3	l_4	$l_{(3+4)}$	
10/17/2012	13:30:17.0	252	1900	0.133	252	3800	0.066	6.1	6.1	12.2	0.199
10/17/2012	13:31:57.0	0	1900	0.000	329	3800	0.086	0.0	6.1	6.1	0.086
10/17/2012	13:33:36.3	142	1900	0.075	142	3800	0.037	6.1	6.1	12.2	0.112
10/17/2012	13:35:17.0	236	1900	0.124	354	3800	0.093	6.1	6.1	12.2	0.217
10/17/2012	13:36:49.6	232	1900	0.122	266	3800	0.070	6.1	6.1	12.2	0.192
10/17/2012	13:38:37.0	144	1900	0.076	216	3800	0.057	6.1	6.1	12.2	0.133

(b) Ring 2, group 2 (phases 7 and 8).											
Date	$t_{BOC,i}$	Phase 1			Phase 2			Lost Time			$(v/s)_{2,2}$
		v_7	s_7	v_7/s_7	v_8	s_8	v_8/s_8	l_7	l_8	$l_{(7+8)}$	
10/17/2012	13:30:17.0	0	1900	0.000	252	3800	0.066	0.0	6.1	6.1	0.066
10/17/2012	13:31:57.0	73	1900	0.038	219	3800	0.058	6.1	6.1	12.2	0.096
10/17/2012	13:33:36.3	0	1900	0.000	249	3800	0.065	0.0	6.1	6.1	0.065
10/17/2012	13:35:17.0	79	1900	0.041	472	3800	0.124	6.1	6.1	12.2	0.165
10/17/2012	13:36:49.6	0	1900	0.000	266	3800	0.070	0.0	6.1	6.1	0.070
10/17/2012	13:38:37.0	0	1900	0.000	324	3800	0.085	0.0	6.1	12.2	0.085

(c) Group 2 critical path determination.										
Date	$t_{BOC,i}$	Compared Values			Lost Times		Selected Values			
		i	$(v/s)_{1,2}$	$(v/s)_{2,2}$	$l_{(3+4)}$	$l_{(7+8)}$	Critical Path	Lost Time	$(v/s)_{G2}$	
10/17/2012	13:30:17.0	512	0.199	0.066	12.2	6.1	34	12.2	0.199	
10/17/2012	13:31:57.0	513	0.086	0.096	6.1	12.2	78	12.2	0.096	
10/17/2012	13:33:36.3	514	0.112	0.065	12.2	6.1	34	12.2	0.112	
10/17/2012	13:35:17.0	515	0.217	0.165	12.2	12.2	34	12.2	0.217	
10/17/2012	13:36:49.6	516	0.192	0.070	12.2	6.1	34	12.2	0.192	
10/17/2012	13:38:37.0	517	0.133	0.085	12.2	12.2	34	12.2	0.133	

TABLE 5.14
 Final steps in the calculation of X_C .

Date	$t_{BOC,i}$	i	$(v/s)_{G1}$	$(v/s)_{G2}$	$\Sigma(v/s)$	l_{G1}	l_{G2}	L	C	$Cl(C-L)$	X_C	CP
10/17/2012	13:30:17.0	512	0.341	0.199	0.54	12.8	12.2	25.0	100.0	1.333	0.720	5634
10/17/2012	13:31:57.0	513	0.173	0.096	0.269	6.4	12.2	18.6	98.6	1.232	0.332	5678
10/17/2012	13:33:36.3	514	0.386	0.112	0.498	12.8	12.2	25.0	101.4	1.327	0.661	5634
10/17/2012	13:35:17.0	515	0.51	0.217	0.727	12.8	12.2	25.0	91.6	1.375	1.000	1234
10/17/2012	13:36:49.6	516	0.256	0.192	0.448	12.8	12.2	25.0	108.4	1.300	0.583	1234
10/17/2012	13:38:37.0	517	0.158	0.133	0.291	6.4	12.2	18.6	100.0	1.229	0.357	5634

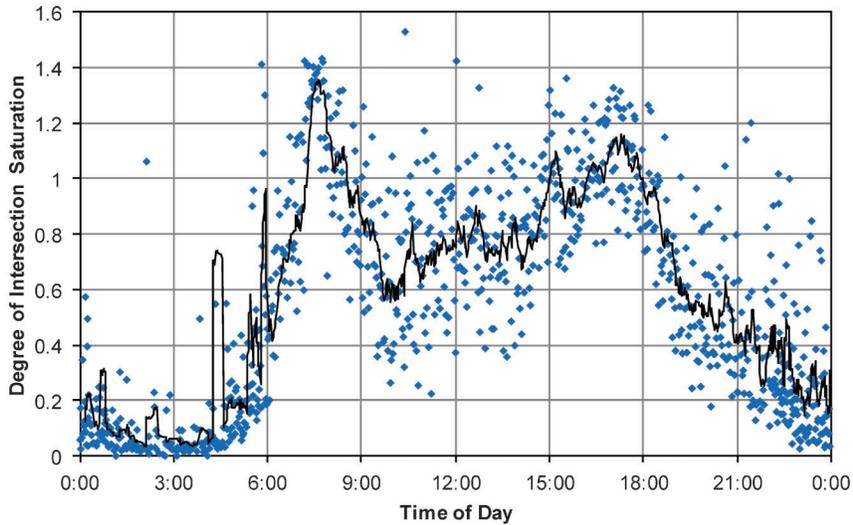


Figure 5.34 Degree of intersection saturation, X_C .

the associated lost time is not included in that cycle's computation of X_C .

Figure 5.34 shows an example plot of X_C for US 36 and Post Road for October 17, 2012. The overall trends in served volumes at the intersection are shown. As perhaps expected, there are two characteristic peaks that correspond to the a.m. and p.m. peak hours. X_C is sustained at above 0.5 more most of the day, with a gradual decrease in the evening and very low volume during the early morning. These are fairly typical trends; a few other interesting small peaks are seen at 15:00 just before the p.m. rush, as well as around 6:00 just before the a.m. rush. The increase around 15:00 likely represents increased traffic to and from schools, and the peak before 6:00 might represent a one-time event of some kind, or it may be caused by a local workplace that opens around that time.

6. PROGRESSION PERFORMANCE MEASURES

The previous chapter presented performance measures oriented toward the measurement of capacity allocation and utilization by the various phases at a signalized intersection and of the entire intersection as a whole. Up to this point, the concept of *signal delay* has not yet been explored. For noncoordinated movements, a random or uniform arrival distribution is typically assumed, and delay estimates are largely a function of the capacity utilization, measured by the volume-to-capacity ratio. Coordinated movements, on the other hand, are sensitive to the characteristics of the arrival profiles, namely, the offset between adjacent signals that controls when platoons of vehicles arrive at the downstream intersection. This chapter introduces performance measures for estimating delay and queue length and for describing the quality of progression through a signalized intersection.

6.1 Vehicle Delay Definitions

Control delay is experienced by travelers whose paths traverse signalized intersections, and is the primary performance measure for signal facilities in the *Highway Capacity Manual* since the 1985 edition. It is therefore worthwhile to spend some time to discuss the meaning of delay and how it relates as a mathematical quantity (both real and theoretical) to vehicle performance on the street. First, let us provide a definition for it.

Control delay is defined as the *increase in travel time* accrued by a vehicle because of traffic control devices, as compared with the travel time if the vehicle were to maintain its expected or desired speed in the absence of the devices.

To define delay, it is necessary to define an expected or desired speed of a vehicle that would be considered to have zero control delay. This is called the *running speed* or *free flow speed*. The measured running speeds of vehicles along any real-world roadway will take on a statistical distribution related to the posted speed limit, driving conditions, and individual driver preferences. In general, the posted speed limit is a reasonable value in many cases.

Figure 6.1 illustrates the definition of control delay for a single vehicle moving through a signalized intersection. Here, a vehicle approaches the intersection while moving at the running speed u_R . At some time, the driver perceives some sort of downstream impedance at the intersection (perhaps a red light, a queue, or both). He or she reacts to those conditions at time t_1 , when vehicle deceleration starts. The vehicle crosses an advance detector at t_D . The vehicle comes to a complete stop at t_2 . At t_3 , the driver begins to accelerate again, crossing the stop bar at t_4 , and finally returning to the

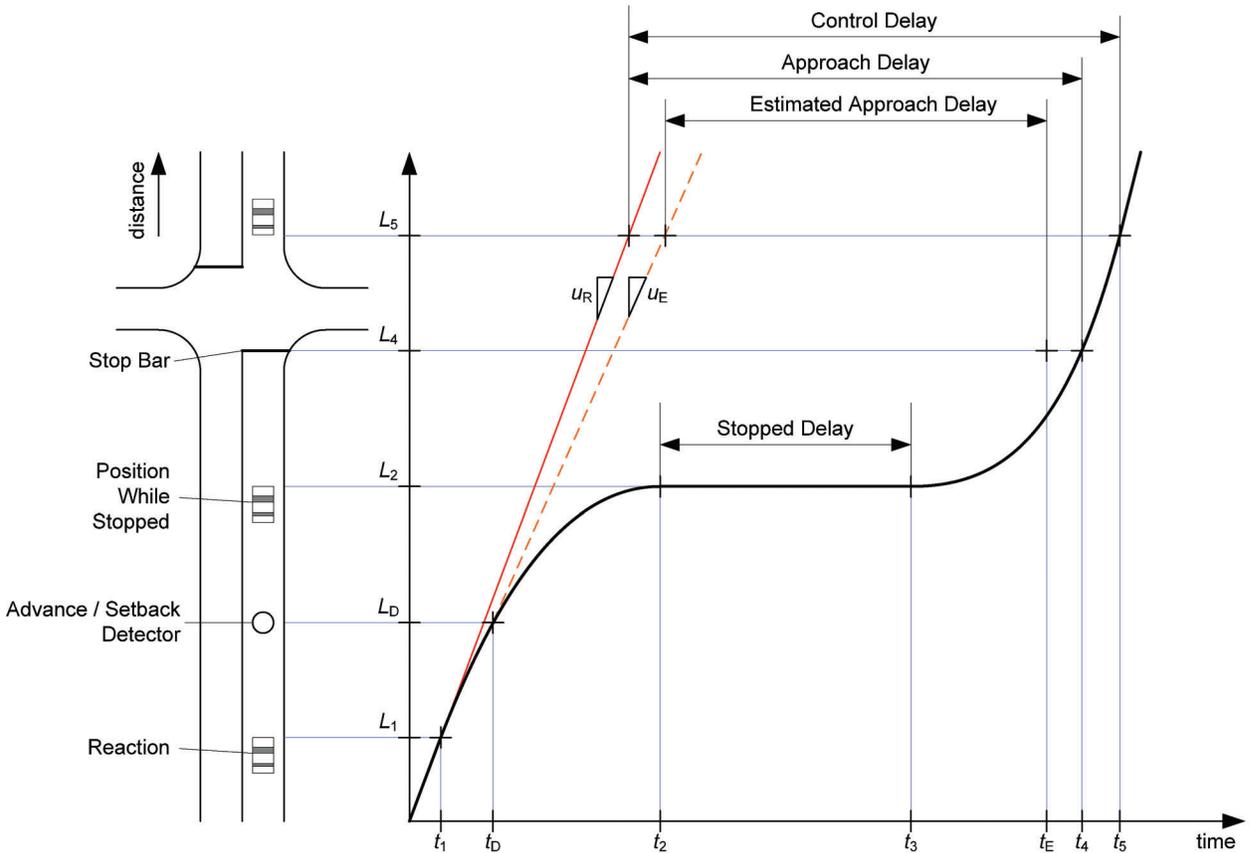


Figure 6.1 Delay definitions, with reference to the perspective of an advanced detector.

running speed at t_5 . The vehicle positions where these events occur are indicated as L_1 , L_2 , etc. The time t_E represents the estimated time when the vehicle is estimated to pass the stop bar. Two lines are drawn tangent to the vehicle trajectory. Tangent to t_1 , the red line shows the ideal vehicle trajectory if it had been able to continue at speed u_R from that time. Tangent to t_D , the orange dashed line shows the estimated ideal vehicle trajectory based on an estimated speed u_E .

The following definitions of delay are illustrated in Figure 6.1:

- *Stopped delay* is defined as the total amount of time that the vehicle is stopped, or simply $t_3 - t_2$.
- The *control delay* is defined as the duration of time between when the vehicle first started to decelerate and when it later finished accelerating, minus the travel time between the two positions, or $t_5 - t_1 - \tau$, where $\tau = (L_5 - L_1)/u_R$.
- The *approach delay* is that portion of control delay that occurs on the approach to the intersection, $t_4 - t_1 - \tau$, where $\tau = (L_4 - L_1)/u_R$.
- The *estimated approach delay* is the value of approach delay based on detector information t_D , the estimated stop bar passage time t_E , and the estimated ideal vehicle speed u_E : $t_E - t_D - \tau$, where $\tau = (L_4 - L_D)/u_E$.

There are several potential sources of error in estimated delay from detector data:

- Approach delay will be less than control delay if vehicle acceleration to u_R continues beyond the stop bar. It has been shown that this difference is about 2 seconds (51), but will probably vary by location.
- Detector times will accurately describe vehicle arrivals if the detector is positioned sufficiently far from the intersection to ensure that $t_D < t_1$ for most vehicles. Detectors too close to the intersection will experience interference from building queues. However, the further upstream the detector, the less likely it is that the vehicle's detection time can accurately predict its arrival time at the intersection (i.e., it is less likely to travel at u_R for the entire distance).
- The selection of a model for estimating vehicle departures, or the times that vehicles cross the stop bar t_E , has an impact on the accuracy of approach delay estimates. Conventional point detectors are unable to track vehicles from arrival to departure, but only provide arrival and possibly departure rates.

6.2 Delay and Quality of Progression

For *groups of vehicles*, the control delay can be estimated by finding the area between the cumulative arrival and cumulative departure curves. The concept is illustrated in Figure 6.2, which contains two related diagrams. Figure 6.2a is a time-space diagram showing the impact of a blockage that occurs at position L_0 on

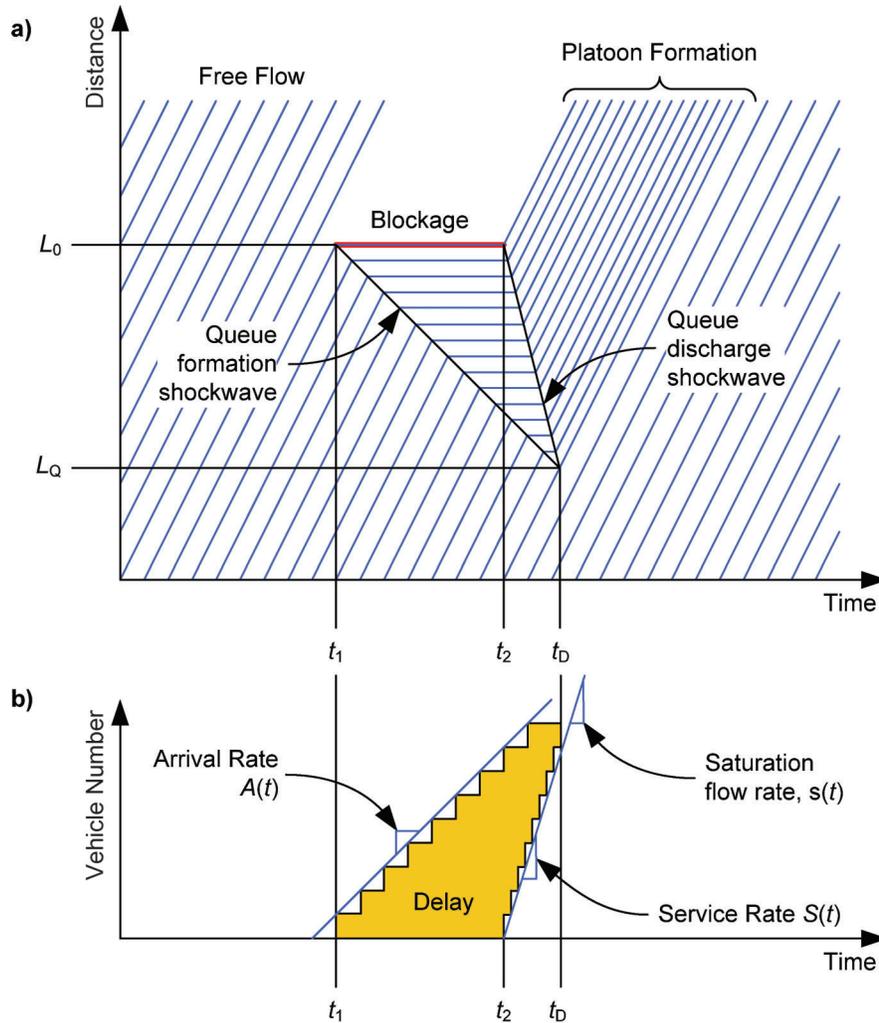


Figure 6.2 Delay and traffic flow characteristics with uniform arrivals: (a) time-space diagram; (b) queue profile.

uniformly distributed traffic flow. Vehicles arriving during the blockage (starting at time t_1) form a queue, which spills backwards as it grows in a shockwave. After the blockage ends (t_2), queued vehicles begin to move again. This creates a second discharge shockwave as the front of the queue begins to clear. When the two shockwaves meet, the queue is fully dispersed, at time t_D . The maximum distance that the back of the queue reaches is L_Q . Vehicles departing from the front of the queue are presumed to move forward at the saturation flow rate, s . Note that the blockage forms a platoon in the otherwise uniformly distributed traffic flow.

Figure 6.2b shows a profile of the queue polygon, showing the arrival and departure time by vehicle number. The number of queued vehicles begins to accrue after t_1 . The vehicles in the front of the queue begin departing at t_2 , and the last vehicle to have been stopped by the blockage begins moving at t_D . The arrival rate $A(t)$ and number of served vehicles $S(t)$ are indicated in this plot; the area between the two lines is equal to the total delay incurred by the vehicles as a

group. Under random arrivals, the arrival rate $A(t)$ is constant, whereas the service rate $S(t)$ is time-dependent (i.e., it is zero when the blockage is present and s when the blockage is absent). If the abstract “blockage” is caused by a traffic signal red time, then the total delay can be directly related to the duration of the red time, which is related to the ratio of green time to cycle length, g/C .

There are several formulas for computing delay at an isolated intersection with random arrivals; Webster proposed what has been one of the most influential (39):

$$d_w = \frac{C(1-g/C)^2}{2[(1-xg/C)]} + \frac{x^2}{2q(1-x)} - 0.65 \left(\frac{C}{v^2}\right)^{1/3} x^{2+5g/C} . \quad (6.1)$$

In this formula, the first term represents the uniform delay, which is expected to be incurred when vehicles arrive with a uniform distribution of headways as in

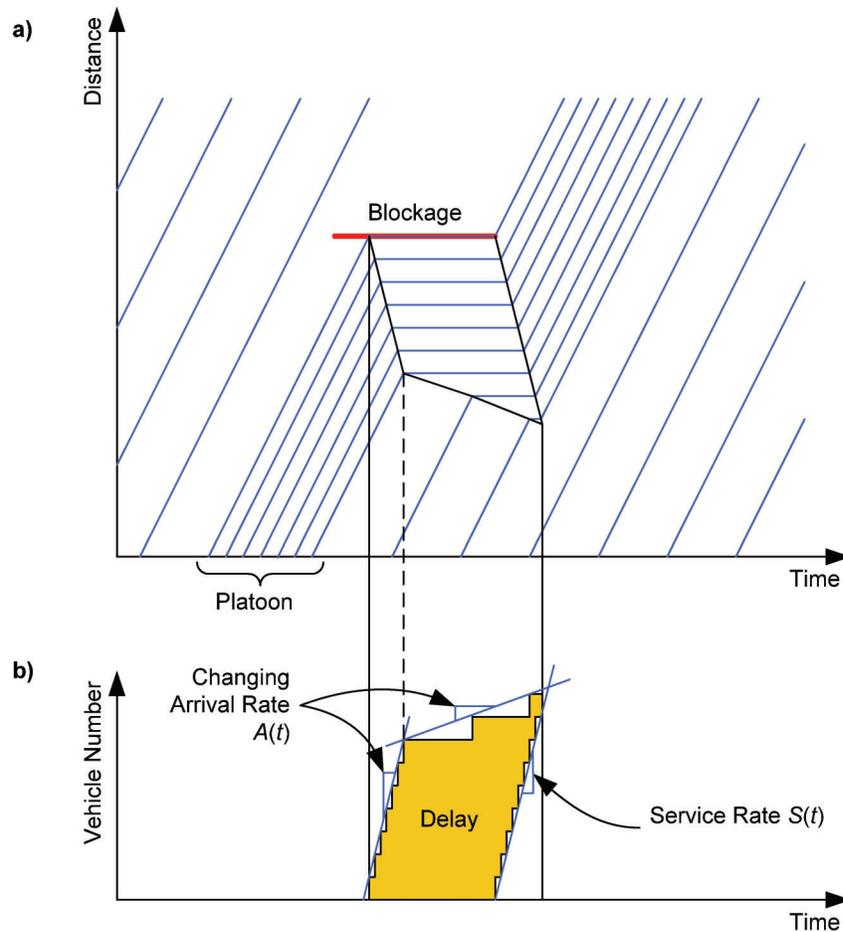


Figure 6.3 Delay and traffic flow characteristics with platoon arrivals: (a) time-space diagram; (b) queue profile.

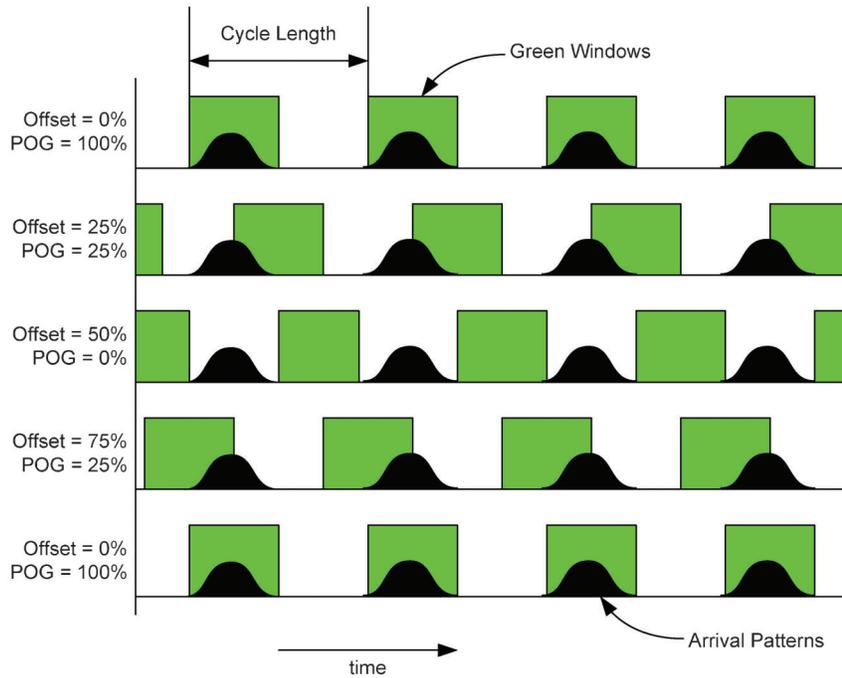
Figure 6.2a. The second term represents the impact of the random nature of vehicle arrivals, which is based upon the assumption of Poisson-distributed vehicle headways and a constant departure rate. The third term is a correction based on findings from simulation.

Of course, many signalized intersections have one or more movements where vehicles do not arrive randomly, but instead appear in platoons formed at upstream intersections. Figure 6.3 illustrates this scenario with a time-space diagram (Figure 6.3a) and the delay polygon (Figure 6.3b). As indicated in Figure 6.3b, $A(t)$ changes at various times, and the leading edge of the delay polygon takes its shape accordingly. Had the blockage avoided the platoon by ending earlier or starting later, there would have been less delay; traffic signals are coordinated in an attempt to achieve this. The analysis of delay in this case is somewhat more complicated given the less straightforward shape of $A(t)$. When a fixed cycle length is used, $A(t)$ and $S(t)$ become oscillatory functions with a period equal to the cycle length C . To minimize delay, the objective is to make the arrival distribution with the bulk of the service time distribution (52). This is primarily achieved by optimizing the offsets of adjacent signals. The arrival pattern $A(t)$ is controlled by the offset of the upstream intersection, and the service

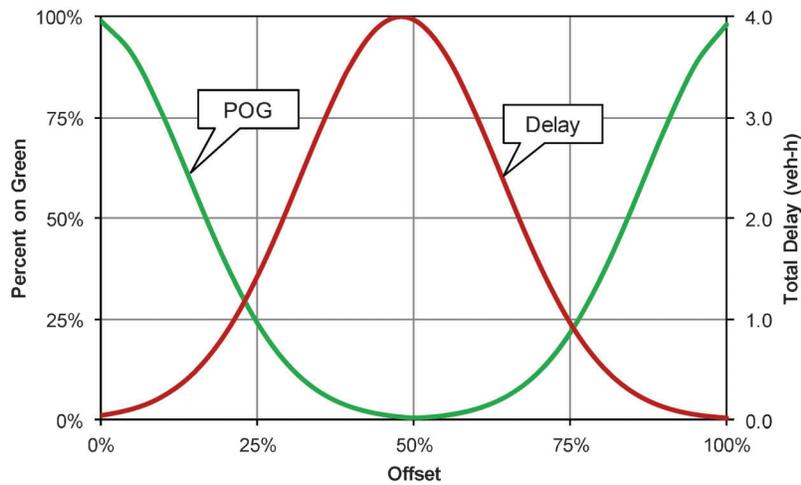
pattern $S(t)$ depends on the offset of the downstream intersection. The duration of the service time is equal to the length of the green time g , and the duration of the blockage or red time is equal to $C - g$.

Figure 6.4 illustrates the development of an offset-performance measure curve. In Figure 6.4a, we see the timelines of vehicle arrival and service (green window) times under several different potential offsets, which are shown in terms of the percentage of the cycle length. Note that 0% and 100% are equivalent; this happens to be the offset where the green window coincides with the arrival pattern the most. The 50% offset causes vehicles to arrive during red; intermediate offsets have intermediate percentages of arrivals during green. Figure 6.4b shows the associated performance curves for this scenario; the minimum delay and maximum percent on green (POG) occur at offset 0%; maximum delay and minimum POG occur at offset 50%. Figure 6.4 conceptually illustrates the strong relationship between delay and the quality of progression (as measured by POG). Measures of progression quality are relevant to understanding the delay performance of a coordinated movement.

In this example, the optimal offset is clearly 0%. However, in most real-world situations, arrival patterns are rarely as clean as those seen in this example;



(a) Conceptual impact of offset on coincidence of vehicle arrivals and green windows



(b) Delay-offset curve

Figure 6.4 Relationship between percent on green, delay, and offset.

secondary platoons also appear in the arrival profiles, as will be seen later in this chapter. Finally, for purposes of optimization, the benefit to one approach must be balanced against the cost to other approaches at the same intersection and its neighboring intersections. Finally, there is some interplay between the sequence of phases and the portion of the cycle time that the green windows occupy. In systems containing a number of intersections the complexity of these interactions can be challenging to perceive, especially during the course of a day, thus necessitating performance measures and other

visualization tools for understanding the quality of progression.

6.3 Progression Performance Measures

The HCM delay equation (40) is

$$d_{\text{HCM}} = (\text{PF})d_1 + d_2 + d_3, \quad (6.2)$$

where d_1 is the uniform delay, d_2 is the incremental delay, and d_3 is the initial queue delay. The uniform

delay is modified by a quantity known as the progression factor (PF), which accounts for arrivals in red or arrivals in green due to the effects of vehicle platoons.

The uniform delay is given by

$$d_1 = \frac{0.5C(1-g/C)^2}{1 - [\min(1, x)g/C]}, \quad (6.3)$$

which is equivalent to the first term of Webster's formula (Equation 6.1). The progression factor is given by

$$PF = \frac{1-P}{1-g/C} f_{PA}, \quad (6.4)$$

where P is the proportion of vehicles arriving on green (or simply "percent on green," POG), f_{PA} is an adjustment factor for platoon arrivals during green (default value 1.0), and the other terms are as defined earlier. As can be seen in this formula, P is the critical component of PF: The greater the value of P , the smaller the value of PF and hence the smaller the value of uniform delay. In the extreme case of $P = 1$, then the uniform delay would be equal to zero.

P is calculated by

$$P = \frac{N_r}{N_r + N_g} = \frac{N_r}{N}, \quad (6.5)$$

where N_r is the number of vehicles arriving during red, N_g the number during green, and N the total for the entire cycle (consisting of one preceding red indication and the subsequent green indication).

The percent of vehicles arriving during green is a basic performance measure to describe vehicle arrivals, but it is correlated very strongly with the balance of phase times at the intersection. If a signal indication for a particular movement is green most of the time, there will obviously tend to be more arrivals on green, regardless of the arrival pattern. It is possible to correct for these effects by dividing P by the g/C ratio, which leads to a quantity known as the "Platoon Ratio" R_p (40):

$$R_p = \frac{P}{g/C} = \frac{CP}{g}. \quad (6.6)$$

The greater the value of R_p , the better the quality of progression.

In addition to PF and R_p , which are related, the HCM defines a performance measure known as the arrival type (AT) to designate six qualitative categories of progression. The corresponding ATs for different values of R_p are described in Table 6.1. This is the equivalent of Exhibit 15-4 from the 2000 HCM, with the addition of an extra column for interpolated values of AT.

Table 6.2 presents example data for POG and AT for phase 2 at US 36 and Post Road on October 17, 2012. Five of the six cycles have a POG higher than 75% ($P_{2,i} > 0.75$). Note that the g/C ratio tends to vary from one cycle to another. In three cycles, it is around 45%, and in three others it is approximately 63%. Thus, while the POG is typically quite high, the arrival type tends to fluctuate between 4 and 6 from one cycle to another. In general, however, the quality of progression on this particular approach is relatively good.

Figure 6.5 shows a plot of phase 2 POG and Figure 6.6 shows a plot of phase 2 AT for the same data set. The signal is coordinated from 6:00 to 22:00, and during almost all of those cycles the POG is above 0.5 and the AT ranks 4 or higher. In general, the quality of progression is satisfactory, with some time periods being better served than others. The a.m. and p.m. peaks have lower POG, which might be caused by increased traffic, or increased secondary platoons from traffic entering the approach from a side-street movement. The late evening/early morning time periods (22:00–24:00 and 0:00–6:00) have greater variation in both POG and AT because the signal is not coordinated during those periods. The intersection runs in a fully actuated "free" mode, and the controller rests in the coordinated phase green in the absence of conflicting phase demand. There are some cycles for which POG is 100%, but the AT is lower than 6. Some arrival types as low as 2 are observed during this time period, which likely reflects either cycles with extremely large g/C ratios, low POG, or both.

TABLE 6.1.
Arrival type definition table, based on HCM Exhibit 15-4 (40).

Arrival Type	Range of Platoon Ratio	Default Value of R_p	Progression Quality	Interpolated Arrival Type Equation
1	$R_p \leq 0.50$	0.333	Very poor	$2R_p + 1$
2	$0.50 < R_p \leq 0.85$	0.667	Unfavorable	$\frac{R_p}{0.35} + \left(3 - \frac{0.85}{0.35}\right)$
3	$0.85 < R_p \leq 1.15$	1.000	Random arrivals	$\frac{R_p}{0.3} + \left(4 - \frac{1.15}{0.3}\right)$
4	$1.15 < R_p \leq 1.50$	1.333	Favorable	$\frac{R_p}{0.35} + \left(5 - \frac{1.5}{0.35}\right)$
5	$1.50 < R_p \leq 2.00$	1.667	Highly favorable	$2R_p + 2$
6	$R_p > 2.00$	2.000	Exceptional	6

TABLE 6.2
Example data table for POG and AT calculation.

Date	$t_{BOC,i}$	i	N_g	N	$P_{2,i}$	C_i	$g_{2,i}$	$g_{2,i}C_i$	AT
10/17/2012	13:30:17.0	512	19	23	0.826	100.0	44.0	0.44	5.75
10/17/2012	13:31:57.0	513	16	19	0.842	98.6	62.3	0.63	4.52
10/17/2012	13:33:36.3	514	32	41	0.780	101.4	63.3	0.62	4.29
10/17/2012	13:35:17.0	515	19	23	0.826	91.6	41.0	0.45	5.69
10/17/2012	13:36:49.6	516	18	26	0.692	108.4	47.7	0.44	5.15
10/17/2012	13:38:37.0	517	18	22	0.818	100.0	62.8	0.63	4.44

6.4 Delay Estimates from Measured Arrival Profiles

At locations where the arrival profiles can be directly measured, it is possible to analyze the delay by calculating the area between the arrival and departure curves (53). This is done by directly considering the cumulative arrivals and departures over time based on vehicle detections and phase status. The departure curve can be measured either directly using departing vehicle counts or by assuming a departure profile based upon the actual green times. This report focuses on the latter case, because there are very few existing intersections that feature both advance detectors and stop bar detectors.

Figure 6.7 illustrates how arrival and departure curves (Figure 6.2 and Figure 6.3) can be obtained from field data. In Figure 6.7a, the intersection arrivals are determined from advance detector counts. Each count is adjusted by a travel time to account for travel time between the detector and the intersection (stop bar). Figure 6.7b shows how a departure curve may be created by assuming a deterministic outflow (at the saturation flow rate) that is approximated as a linear function beginning after the start of effective green and terminating after the departure of the queue or the end of the green time, whichever comes first.

Conceptually, the area between the arrival and departure curves is equivalent to the following (52):

$$d = \int_{t_0}^{t_0+T} [q(t_0) + A(t) - D(t)] dt, \quad (6.7)$$

where

$q(t_0)$ = queue length at time t_0 (number of vehicles);

$A(t)$ = arrival rate (vehicles per unit time);

$D(t)$ = service, or departure, rate (vehicles per unit time);

t_0 = beginning of analysis period;

T = duration of analysis period.

In the actual data, of course, $A(t)$ and $D(t)$ are not functions but are defined by discrete events: the vehicle arrival times $t_{A,k}$, the beginning of green t_g , and the end of green t_r . The analysis period comprises one service instance spanning two consecutive ends of green. With these events, we can divide the continuous timeline into discrete chunks. For this analysis, the events are numbered using the index $k = 1, 2, \dots$, where the k th interval begins at t_{k-1} and ends at t_k . At every event time t_k , there is a certain queue length q_k that is attained at that particular time (the end of the interval).

The queue length for each event time can be found as follows (30):

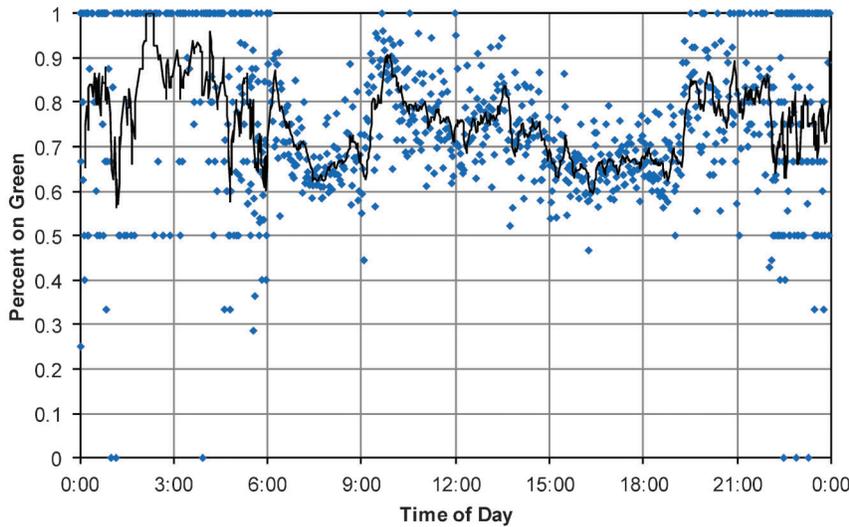


Figure 6.5 Percent on green.

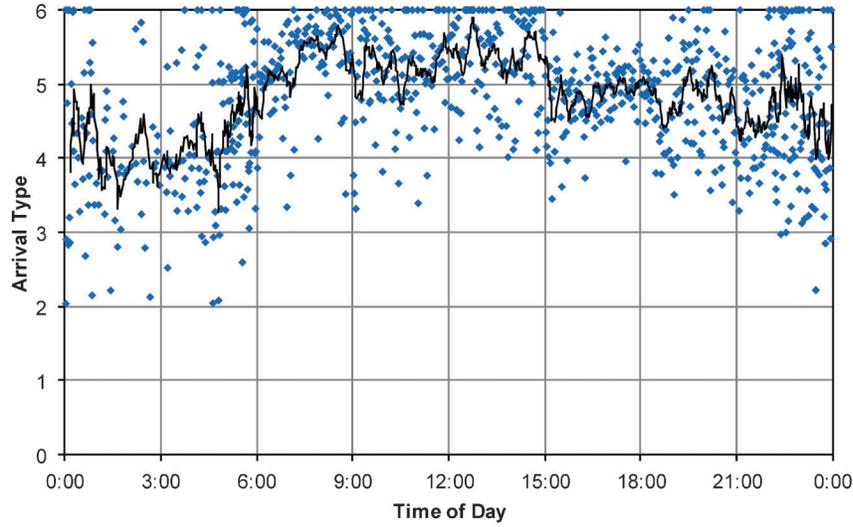


Figure 6.6 Arrival type.

if $t_k \leq t_g$ (effective red),

$$q_k = \begin{cases} q_{k-1} + 1 & \text{if event type is vehicle arrival,} \\ q_{k-1} & \text{if event type is beginning of green;} \end{cases}$$

If $t_k > t_g$ (effective green),

$$q_k = \begin{cases} \max(0, q_{k-1} - c_k + 1) & \text{if event type is vehicle arrival,} \\ \max(0, q_{k-1} - c_k) & \text{if event type is end of green,} \\ 0 & \text{if } c_{k-1} \geq q_{k-1}. \end{cases} \quad (6.8)$$

The interval capacity c_k is defined as

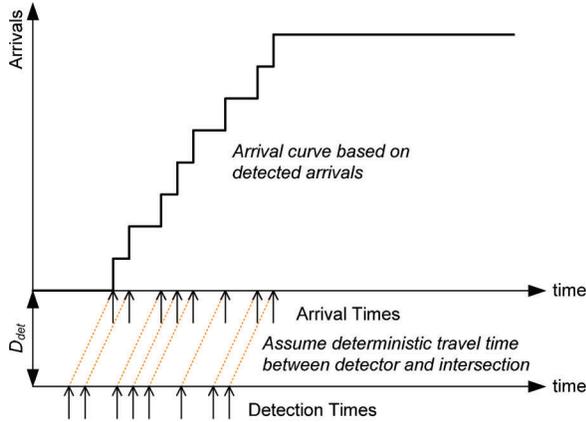
$$c_k = \begin{cases} 0 & \text{if } t_k \leq t_g, \\ s(t_k - t_{k-1}) & \text{if } t_k > t_g, \end{cases} \quad (6.9)$$

where s is the saturation flow rate in vehicles per second. The delay that accrues during each interval is calculated from (30,53)

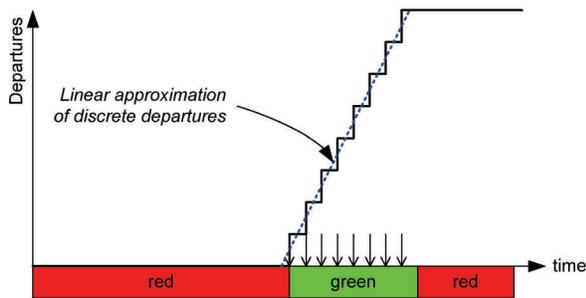
$$d_k = \begin{cases} q_{k-1}(t_k - t_{k-1}) & \text{(a) if } t_k \leq t_g, \\ (q_{k-1} - \frac{1}{2}c_k)(t_k - t_{k-1}) & \text{(b) if } t_k > t_g \text{ and } c_k < q_{k-1}, \\ \frac{1}{2}q_{k-1}(t_D - t_{k-1}) & \text{(c) if } t_k > t_g \text{ and } c_k \geq q_{k-1}. \end{cases} \quad (6.10)$$

The meaning of the three above scenarios is explained below, with corresponding graphics in Figure 6.8.

- (a) The rectangular area (Figure 6.8a) models the growth of the queue during red. During each interval, the queue is considered to be of length q_{k-1} . At the end of the interval, the queue grows to length $(q_{k-1} + 1)$.
- (b) The trapezoidal area (Figure 6.8b) models an interval in which the queue size is decreasing but does not discharge completely before the next event. During the interval, the queue reduces linearly by the total amount c_k .

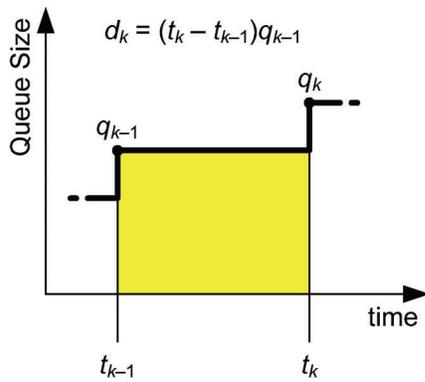


(a) Arrival profile.

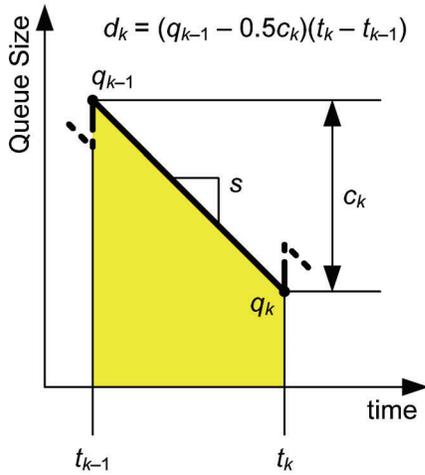


(b) Departure profile.

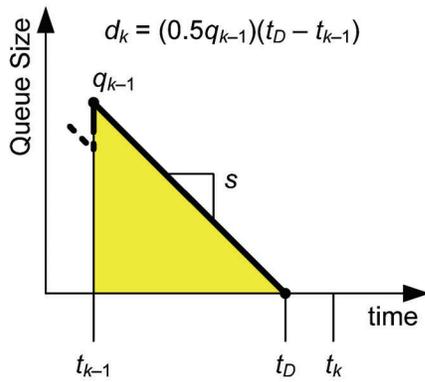
Figure 6.7 Obtaining arrival and departure profiles from field data.



(a) Rectangular



(b) Trapezoidal



(c) Triangular

Figure 6.8 Input-output delay polygons (30).

- (c) The triangular area (Figure 6.8c) models an interval in which the queue fully discharges before the end of the interval ($c_k \geq q_{k-1}$). The discharge time t_D is calculated as

$$t_D = q_{k-1}/s. \quad (6.11)$$

Example arrival and departure curves for a typical signal cycle and the resulting queue profile are shown in Figure 6.9. The arrival curve (Figure 6.9a) is determined by the discrete arrival events obtained from

actuators of the setback detector, adjusted for travel time to the intersection (Figure 6.7a). The departure curve (Figure 6.9b) is a linear function of the saturation flow rate s that begins from time t_g , the beginning of the effective green. The combination of these two curves yields the queue profile (Figure 6.9c), subject to the rule that $q_k \geq 0$. The area under this curve is the total delay.

Table 6.3 presents example data from two cycles. Each line in this table represents an event, labeled as a vehicle arrival, beginning of green (BOG), or end of green (EOG). The interval “width” $T_k = t_k - t_{k-1}$ is given, and the corresponding capacity c_k . For each interval, the discharge time T_D is computed, which tells the time needed to clear the queue present at time t_{k-1} . During green, T_D determines the triangular delay area when $c_k \geq q_k$, or $T_D \leq T_k$ (Figure 6.8c). D_k is the number of discharged vehicles. This is equal to the capacity when $c_k < q_k$ (the queue has not yet cleared). $D_k = 1$ when the event is a vehicle arrival and the queue has cleared. Vehicles arriving in green after queue clearance do not contribute any further delay. The delay associated with each interval, d_k , results from the application of Equation 6.10. The total delay accrued, d_{total} , is the sum of d_k across all intervals. The average delay d_{avg} is found at the EOG by dividing d_{total} by N (the total number of arrivals). A plot of the queue profile and the number of vehicles in queue corresponding to these data is included in Figure 6.10.

Figure 6.11 shows a plot of the estimated average delay for phase 2 (westbound through movement) at US 36 and Post Road on October 17, 2012. The average delay thresholds for the HCM level of service (LOS) for signalized intersections are also included in the figure. During the 24-hour period, there are no cycles where the average delay extends to the region of LOS F. The worst delays are experienced in the a.m. peak, which is unsurprising given that the morning has the highest volumes for the westbound through movement (see Figure 5.20). During the rest of the day, the average delay is usually in the LOS B range, with some cycles having higher or lower average delay. With the exception of several cycles just after midnight, the overnight period tends to have very low average delay for the through movement.

6.5 Purdue Coordination Diagram

Aside from the generation of performance measures, the event data also supports visualization of the quality of progression using a useful graphic called the Purdue Coordination Diagram (PCD). This graphical tool has been described extensively in several previous papers (54–56).

A coordination diagram for one signal cycle is shown in Figure 6.12. The two axes of the plot are both time: The vertical axis is the time in cycle and the horizontal axis is the time of day. The cycle is defined by two intervals: a preceding effective red and a subsequent effective green. Together, these two intervals add up to the cycle length. These two intervals are defined by

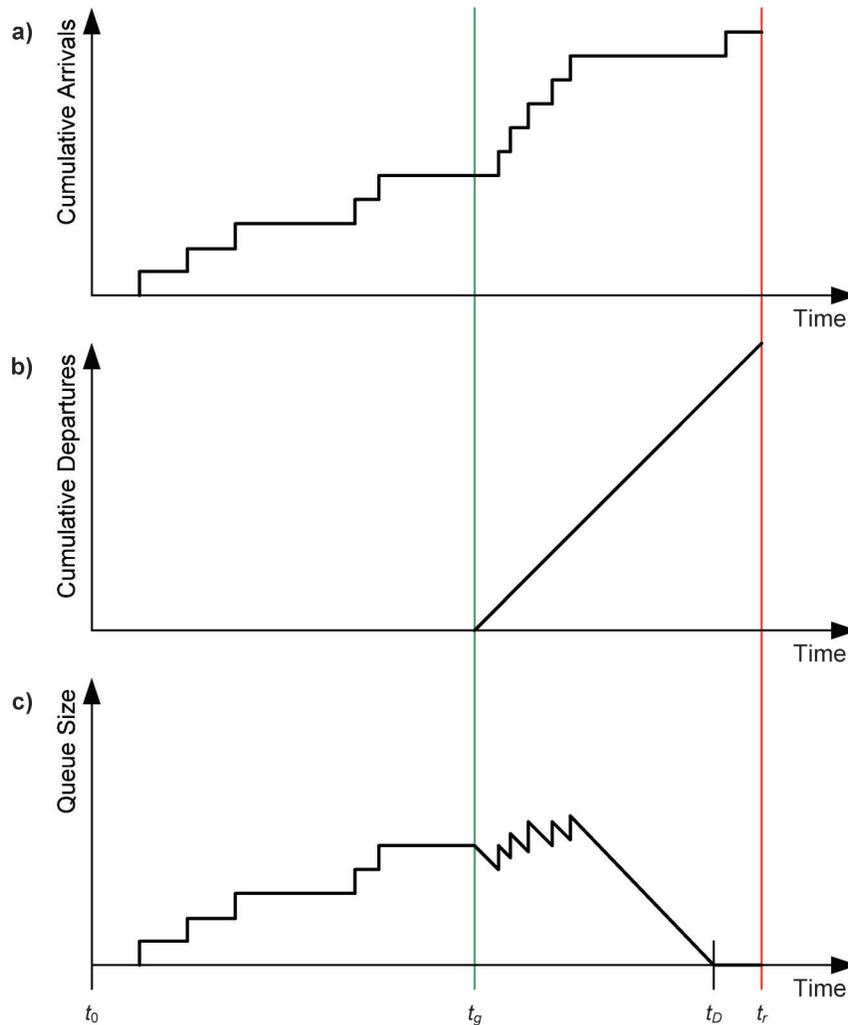


Figure 6.9 Concepts for “input-output” delay estimation (30).

three times: the last end of green (LEOG), beginning of green (BOG), and the (current) end of green (EOG). This defines a rectangular area, the diagonal of which is the meaning of “now” in the time-time plot; vehicle arrivals (detector actuations adjusted for travel time to the stop bar) are plotted as dots that fall on this line, with the arrival time of each vehicle $t_{A,k}$ located according to the time of day and time in cycle. This yields an event view that visually displays the progression situation for that particular cycle.

Although this might not be a particularly informative view for one cycle, when multiple cycles are shown in succession, patterns begin to emerge. Figure 6.13 shows a coordination diagram for a 30-minute period. The state change events are indicated by markers on the right-hand side of the figure. Consistent arrival patterns are seen from one cycle to another—with some stochastic variations throughout the time period. In nearly every cycle, two distinct platoons of vehicles can be observed: one originating from the upstream coordinated movement (Figure 6.13, callout “i”) and another secondary platoon that corresponds to vehicles entering from the side-street (Figure 6.13, callout “ii”).

Because the signal operates with a fixed cycle length, subsequent cycles are vertically aligned with excellent agreement. There is a considerable amount of variation in the BOG caused by phase actuation.

When expanded to a 24-hour view, it becomes possible to see whether these trends hold throughout the day, and how they vary. Figure 6.14 shows an expanded view of the operation of northbound phase 6 from the intersection of SR 37 and Greenfield Avenue in Noblesville, Indiana, on July 25, 2009. This plot was selected because of the strong and clear appearance of primary and secondary platoons throughout the day. Some variation in the beginning of green can also be seen. Minor variations in the end of green appear because of the actuation of the end of coordinated green (see also Figure 3.17).

Figure 6.15 shows a PCD for 24 hours for westbound phase 2 at US 36 and Post Road on October 17, 2012. This allows a direct comparison with the other example graphics for the same movement provided throughout this document. From this figure it is clear that the heaviest demand occurs during the a.m. peak period (6:00–9:00), with lower demand during the rest of the

TABLE 6.3
 Example calculation of input-output delay for two service instances of phase 2 on October 17, 2012.

k	Event Type	Green	Event Time	T_k	c_k	T_D	D_k	q_{k-1}	q_k	Clear	d_k	d_{total}	N	d_{avg}
1	Vehicle		13:43:34.5	3.9	0.00	0.00	0.00	0.00	1.00		0.00	0.0	1	
2	Vehicle		13:43:35.3	0.8	0.00	0.63	0.00	1.00	2.00		0.80	0.8	2	
3	Vehicle		13:43:36.2	0.9	0.00	1.26	0.00	2.00	3.00		1.80	2.6	3	
4	Vehicle		13:43:36.4	0.2	0.00	1.89	0.00	3.00	4.00		0.60	3.2	4	
5	Vehicle		13:43:38.2	1.8	0.00	2.53	0.00	4.00	5.00		7.20	10.4	5	
6	Vehicle		13:43:40.5	2.3	0.00	3.16	0.00	5.00	6.00		11.50	21.9	6	
7	Vehicle		13:43:43.8	3.3	0.00	3.79	0.00	6.00	7.00		19.80	41.7	7	
8	Vehicle		13:43:53.9	10.1	0.00	4.42	0.00	7.00	8.00		70.70	112.4	8	
9	Vehicle		13:44:37.1	43.2	0.00	5.05	0.00	8.00	9.00		345.60	458.0	9	
10	Vehicle		13:44:41.3	4.2	0.00	5.68	0.00	9.00	10.00		37.80	495.8	10	
11	Vehicle		13:44:43.7	2.4	0.00	6.32	0.00	10.00	11.00		24.00	519.8	11	
12	BOG		13:44:45.0	1.3	0.00	6.95	0.00	11.00	11.00		14.30	534.1	11	
13	Vehicle	✓	13:44:45.3	0.3	0.48	6.95	0.48	10.53	11.53		0.21	534.3	12	
14	Vehicle	✓	13:44:45.8	0.5	0.79	7.28	0.79	10.73	11.73		0.59	534.9	13	
15	Vehicle	✓	13:44:50.4	4.6	7.28	7.41	7.28	4.45	5.45		50.26	585.2	14	
16	Vehicle	✓	13:44:52.5	2.1	3.33	3.44	3.33	2.13	3.13		10.47	595.6	15	
17	Vehicle	✓	13:44:55.8	3.3	5.23	1.97	3.13	0.00	0.00	✓	3.08	598.7	16	
18	Vehicle	✓	13:44:57.0	1.2	1.90	0.63	1.00	0.00	0.00	✓	0.00	598.7	17	
19	Vehicle	✓	13:44:59.0	2.0	3.17	0.63	1.00	0.00	0.00	✓	0.00	598.7	18	
20	EOG	✓	13:45:08.1	9.1	14.41	0.63	1.00	0.00	0.00	✓	0.00	598.7	18	33.26
1	Vehicle		13:45:13.3	5.2	0.00	0.00	0.00	0.00	1.00		0.00	0.0	1	
2	Vehicle		13:45:16.0	2.7	0.00	0.63	0.00	1.00	2.00		2.70	2.7	2	
3	Vehicle		13:45:16.4	0.4	0.00	1.26	0.00	2.00	3.00		0.80	3.5	3	
4	Vehicle		13:45:21.7	5.3	0.00	1.89	0.00	3.00	4.00		15.90	19.4	4	
5	Vehicle		13:45:32.2	10.5	0.00	2.53	0.00	4.00	5.00		42.00	61.4	5	
6	Vehicle		13:45:35.0	2.8	0.00	3.16	0.00	5.00	6.00		14.00	75.4	6	
7	Vehicle		13:45:38.6	3.6	0.00	3.79	0.00	6.00	7.00		21.60	97.0	7	
8	Vehicle		13:45:39.5	0.9	0.00	4.42	0.00	7.00	8.00		6.30	103.3	8	
9	Vehicle		13:45:46.8	7.3	0.00	5.05	0.00	8.00	9.00		58.40	161.7	9	
10	Vehicle		13:45:54.0	7.2	0.00	5.68	0.00	9.00	10.00		64.80	226.5	10	
11	BOG		13:46:11.3	17.3	0.00	6.32	0.00	10.00	10.00		173.00	399.5	10	
12	Vehicle	✓	13:46:14.2	2.9	4.59	6.32	4.59	5.41	6.41		19.97	419.5	11	
13	Vehicle	✓	13:46:26.5	12.3	19.48	4.05	6.41	0.00	0.00	✓	12.97	432.4	12	
14	Vehicle	✓	13:46:29.8	3.3	5.23	0.63	1.00	0.00	0.00	✓	0.00	432.4	13	
15	Vehicle	✓	13:46:30.8	1.0	1.58	0.63	1.00	0.00	0.00	✓	0.00	432.4	14	
16	Vehicle	✓	13:46:32.4	1.6	2.53	0.63	1.00	0.00	0.00	✓	0.00	432.4	15	
17	Vehicle	✓	13:46:33.8	1.4	2.22	0.63	1.00	0.00	0.00	✓	0.00	432.4	16	
18	Vehicle	✓	13:46:34.3	0.5	0.79	0.63	1.00	0.00	0.00	✓	0.00	432.4	17	
19	Vehicle	✓	13:46:37.9	3.6	5.70	0.63	1.00	0.00	0.00	✓	0.00	432.4	18	
20	Vehicle	✓	13:46:38.5	0.6	0.95	0.63	1.00	0.00	0.00	✓	0.00	432.4	19	
21	Vehicle	✓	13:46:38.9	0.4	0.63	0.63	1.00	0.00	0.00	✓	0.00	432.4	20	
22	Vehicle	✓	13:46:40.0	1.1	1.74	0.63	1.00	0.00	0.00	✓	0.00	432.4	21	
23	Vehicle	✓	13:46:44.7	4.7	7.44	0.63	1.00	0.00	0.00	✓	0.00	432.4	22	
24	Vehicle	✓	13:46:47.9	3.2	5.07	0.63	1.00	0.00	0.00	✓	0.00	432.4	23	
25	Vehicle	✓	13:46:48.1	0.2	0.32	0.63	1.00	0.00	0.00	✓	0.00	432.4	24	
26	EOG	✓	13:46:50.6	2.5	3.96	0.63	1.00	0.00	0.00	✓	0.00	432.4	24	18.02

day. During the midday (9:00–15:00), the heaviest bulk of the vehicle distribution appears to arrive during the green window, showing a good quality of progression. The platoons seem to carry over beyond the EOG and into the next cycle, as can be seen by the clusters of dots just above the x -axis (LEOG) during the morning hours. There is also a great deal of early return to green occurring at this time. The p.m. peak period (15:00–19:00) is less clear, with about the same number of vehicles appearing to arrive in red as well as green. There is not an evident primary platoon here either. The

overnight period (0:00–6:00 and 22:00–24:00) sees a huge amount of variation in the end of green and in the duration of the effective cycles because the signal operates in fully actuated mode during those times.

Figure 6.16 shows a PCD that illustrates operation under adaptive control (42). The gradual movements of the vehicle platoons relative to the green bands demonstrate the impacts of online offset tuning during the runtime of a coordination pattern (46). For example, at 10:30, a coordination pattern goes into effect that has vehicles arriving slightly before the start

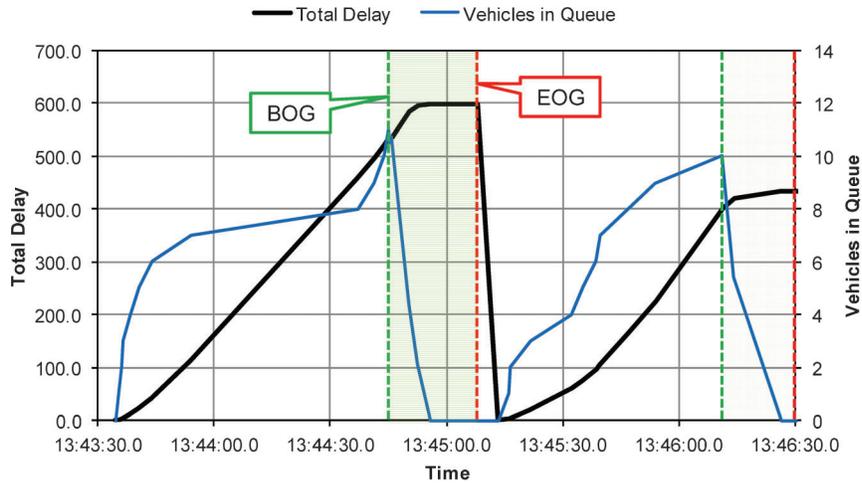


Figure 6.10 Estimated delay (input-output method)—calculation example.

of green (callout “i”), but the offset is incrementally adjusted until the arrivals are positioned in the middle of the green hand (callout “ii”). Similarly, at the beginning of the next pattern at 14:30, the arrivals initially take place during red, with the platoon cut off by the end of green (callout “iii”), but the adaptive offset adjustment eventually brings the platoon arrivals into the green band (callout “iv”). The ability to view the impacts of adaptive decisions is potentially very powerful in evaluating the effectiveness of advanced control methods and assisting in the selection of parameters that govern their behavior. For example, the offset adjustments in Figure 6.16 improve the operation, but there seems to be an opportunity to increase the responsiveness of the algorithm to accelerate the adjustments so that the improvements occur more rapidly. Such information could be valuable to the system operator, the contractor providing the software, or the researchers investigating the algorithms.

Combining vehicle detector information with the status of an *upstream* intersection enables the information

about vehicle arrivals to be enhanced. When the distances between two intersections are relatively short, it may be possible to determine the most likely sources of vehicles at the upstream intersection. Figure 6.17 illustrates the possible sources and destination of vehicles on a link between two intersections. At many signalized intersections, the through and left-turn movements are controlled by specific signal phase or overlap. By looking *backward* from the detection time, to the state of the upstream intersection, it is possible to identify the current phase or overlap in operation.

Figure 6.18 illustrates an application of this concept for a single-controller diamond interchange (57). As illustrated in the figure, vehicles on the southbound link between the two intersections that comprise the diamond interchange originate from either phase 6 or phase 7 at the north intersection, and their destination is overlap D at the south intersection. The time-space diagram shows the location of the detector (callout “i”) and two categories of vehicles detected there:

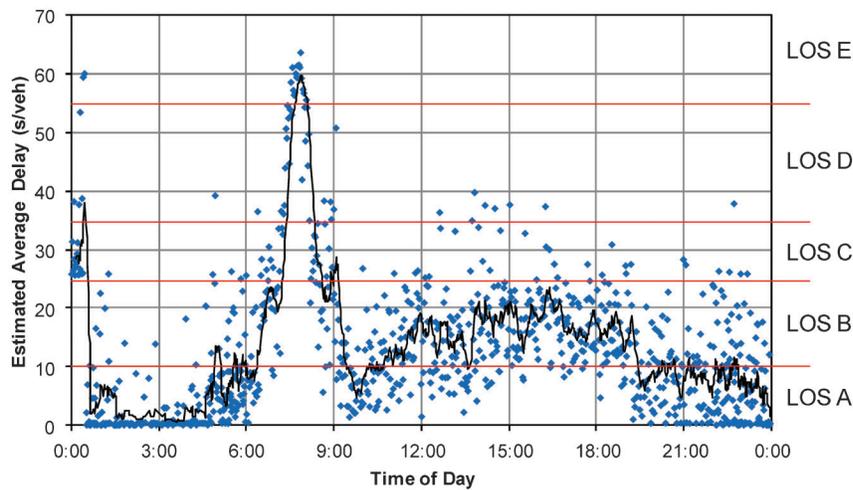


Figure 6.11 Estimated delay (input-output method).

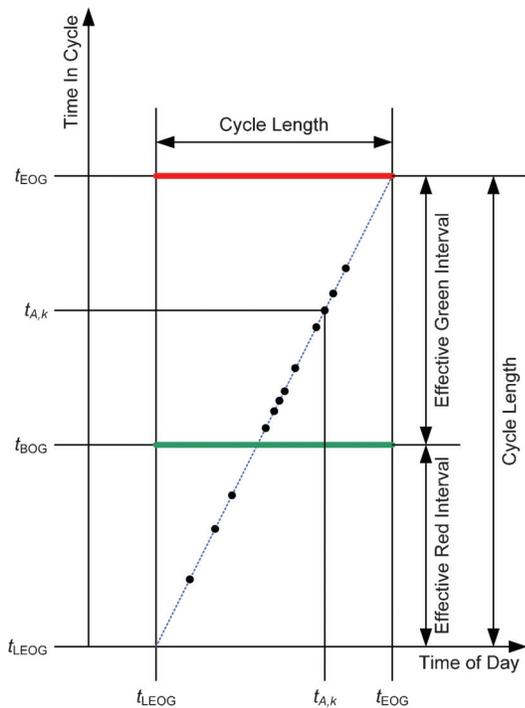


Figure 6.12 Coordination diagram for one cycle.

- One group of vehicles (callout “ii”) is projected forward as arrive during the green for overlap D (callout “iii”), and projected backward as originating from phase 6 (callout “iv”).
- The other group (callout “v”) is projected forward as arrive during the green for overlap D (callout “vi”), and projected backward as originating from phase 6 (callout “vii”).

There is often a gap between the groups (callout “viii”) representative of the lost time between phases 6 and 7. Extending this information to the PCD, it is possible to infer the quality of progression for each specific movement through the intersection pair. Figure 6.19 shows a PCD with each vehicle arrival coded by color according to its most likely source phase. Phase 7 vehicles are represented as gray triangles

(callout “i”), and phase 6 vehicles are the black diamonds (callout “ii”). The impact of the phase sequence can be seen, as it changes according to time of day. During the 09:00–15:00 plan, phase 6 vehicles (callout “iii”) arrive before the phase 7 vehicles (callout “iv”). The PCD also exhibits an *absence* of vehicles during the cycle time in which phase 5 is served. This represents a particularly long share of the cycle time during the 06:00–09:00 plan (callout “v”).

At a diamond interchange, the number of vehicles desiring access to the limited-access facility often represent a greater demand than vehicles trying to pass through both intersections. In Figure 6.19, this would explain the fact that phase 7 vehicles are often greater in number than phase 6 vehicles, especially during the a.m. and p.m. peak periods, and the “empty” portions of the cycle when phase 5 is in service.

6.6 Flow Profiles

The PCD is helpful for developing a quantitative view of signal operations. Another visualization tool is the distribution of vehicles and green times through the cycle. The cyclic flow profile concept was introduced in early studies of vehicle discharge from traffic signals in the 1950s (58) and was a central component in the TRANSYT signal timing optimization software that was originally developed in the late 1960s (59). In TRANSYT, the flow profiles are superimposed with an indication of when a fixed-time green window occurs. The flow profiles represented in this section combine the profile of vehicle arrivals with a distribution of the *probability of green* reflecting actuated operations, rather than a fixed-time green window.

The mathematics to produce a cyclic flow profile depends on the assumption that a fixed cycle length operates during the entire analysis period without interruption. This is typical of normal signal operations, so flow profiles can be developed for most coordinated systems. The timeline is divided into discrete bins, all having the same duration. The resolution r represents the size of each bin. In the discussion provided here, $r = 1$. We then define an analysis period spanning $[t_0, t_0 + T]$, where T is the

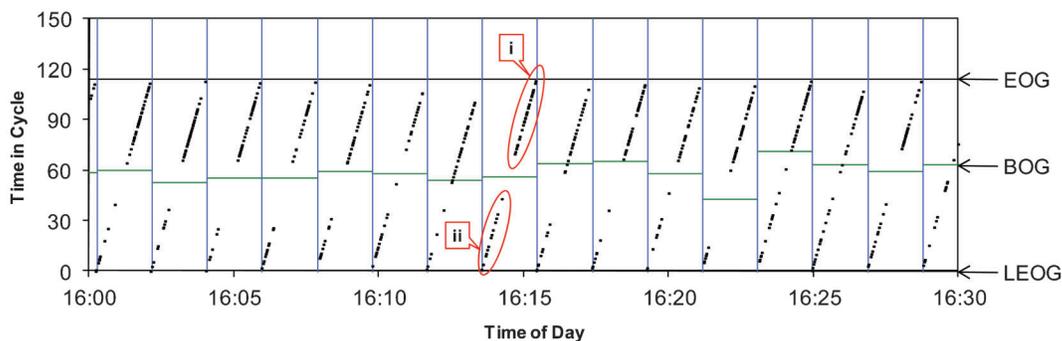


Figure 6.13 Coordination diagram for a 30-minute period.

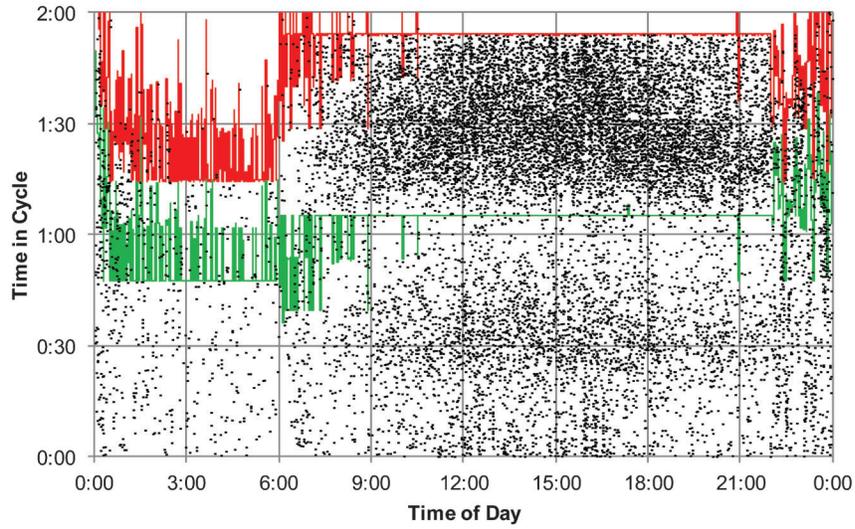


Figure 6.14 Coordination diagram for 24 hours: northbound phase 6, SR 37 and Greenfield Avenue, July 25, 2009.

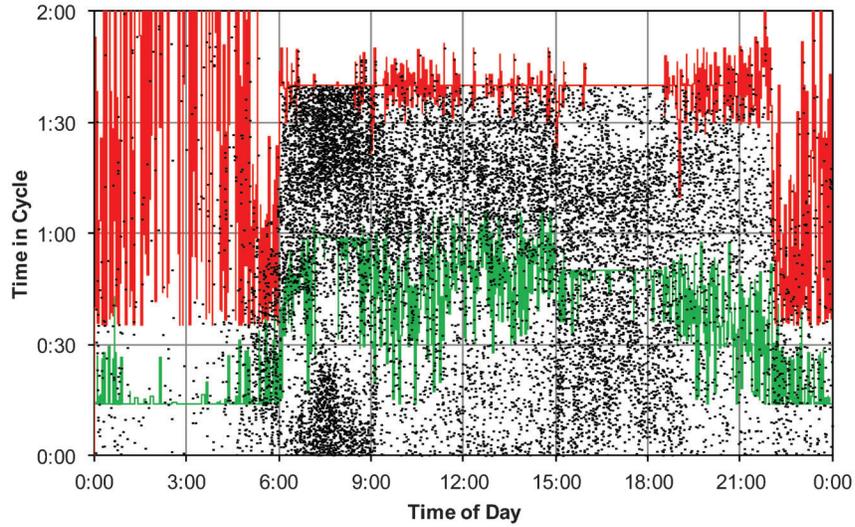


Figure 6.15 Coordination diagram for 24 hours: westbound phase 2, US 36 and Post Road, October 17, 2012.

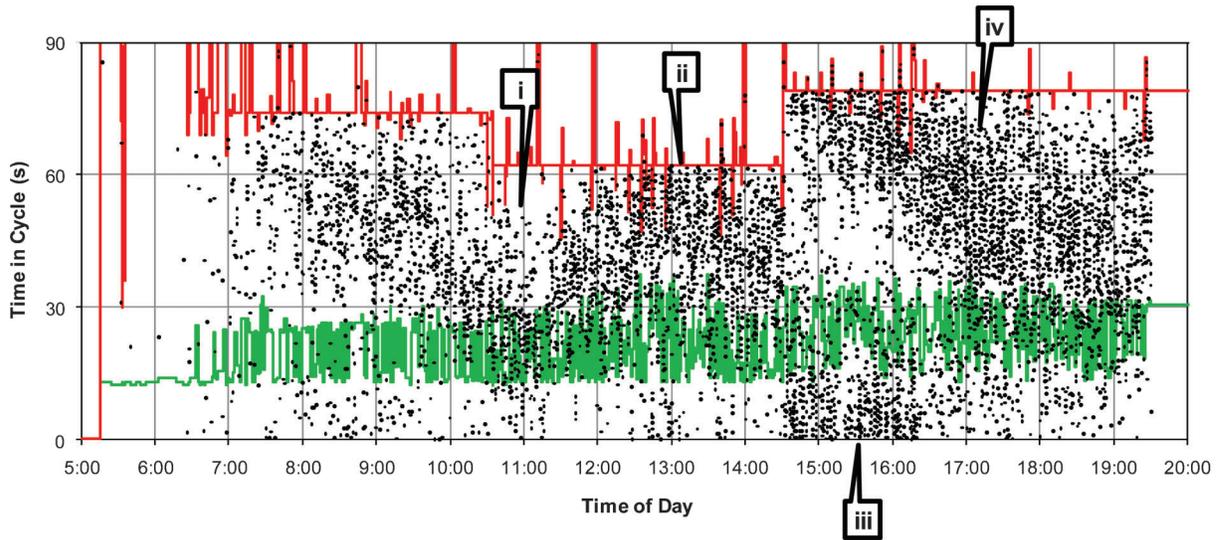


Figure 6.16 Coordination diagram for adaptive control in a simulation environment (42).

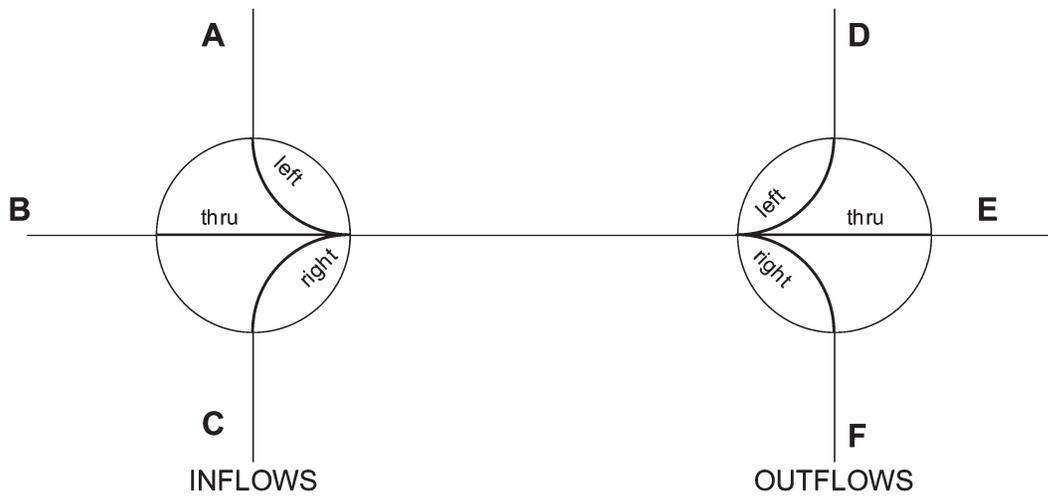


Figure 6.17 Vehicle origins and destinations on a link (7).

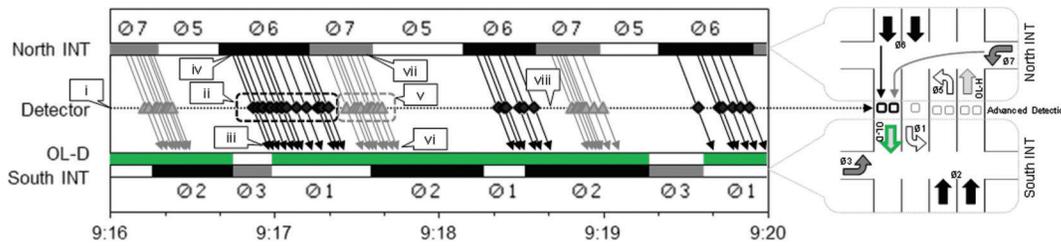


Figure 6.18 Using upstream phase status to determine vehicle origins (57).

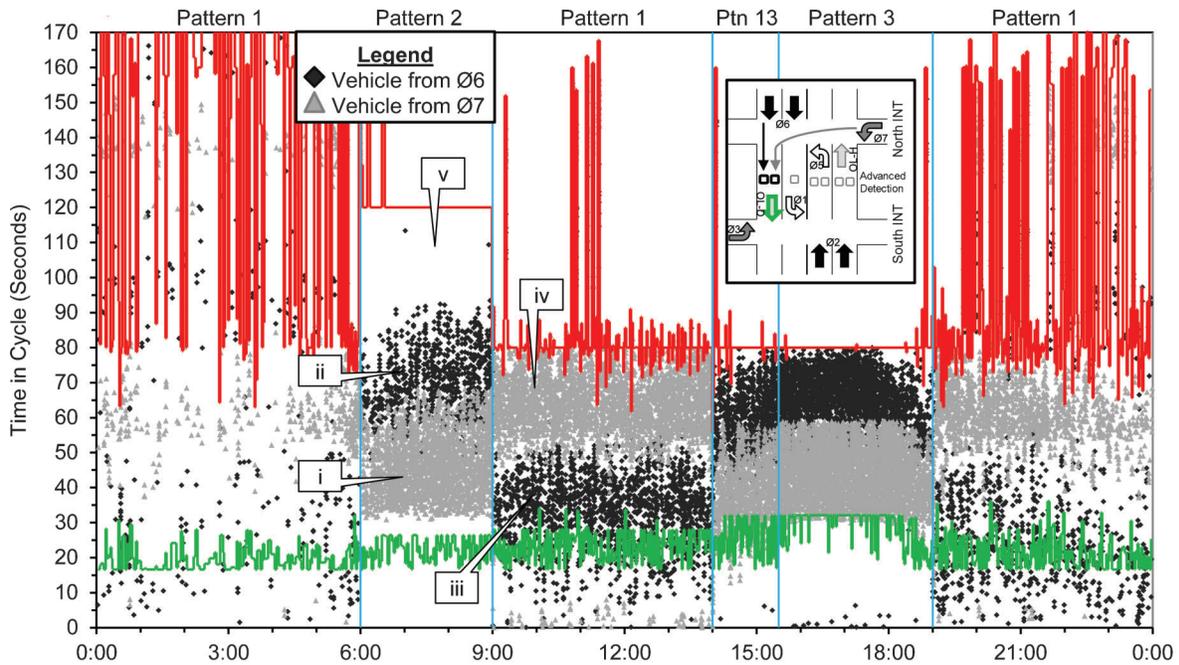


Figure 6.19 Coordination diagram for the southbound through movement at a diamond interchange (I-465 and SR 37 South in Indianapolis, Indiana) for Wednesday, June 5, 2013 (57).

duration of the analysis period and is some multiple of the cycle length C .

The same events that allowed the input-output delay estimate in the previous section now populates two curves: the vehicle arrival distribution $N(t)$ and probability of green distribution $G(t)$, which is related to the capacity. Similar to the production of PCDs, we can aggregate across cycles. Rather than successive ends of greens as the reference points, however, the *system clock* is the frame of reference.

In signal controllers running time base coordination, the system zero occurs at multiples of C beginning from the “pattern sync reference” time (usually midnight by default). If time is expressed as the number of seconds,

for $C = 90$, the system zeroes occur at 0, 90, 180, 270, and so on throughout the rest of the day. Consider a timing plan with $C = 56$ that begins at 13:12:00, or 47,520 seconds after midnight. The number 47,520 is not a multiple of 56; the closest multiple is 47,488. That is taken as the system zero time used as a reference for coordination.

With the cycle reference point in mind, the next step is to aggregate across cycles. The method of doing this is illustrated in Figure 6.20. This shows the timeline of events within $[t_0, t_0 + T]$ being divided into a template matrix comprising $M = C/r$ columns and $Z = T/C$ rows. The columns are numbered as 0, 1, 2, ..., $M - 1$. The numbering of the rows is unimportant. Each cell

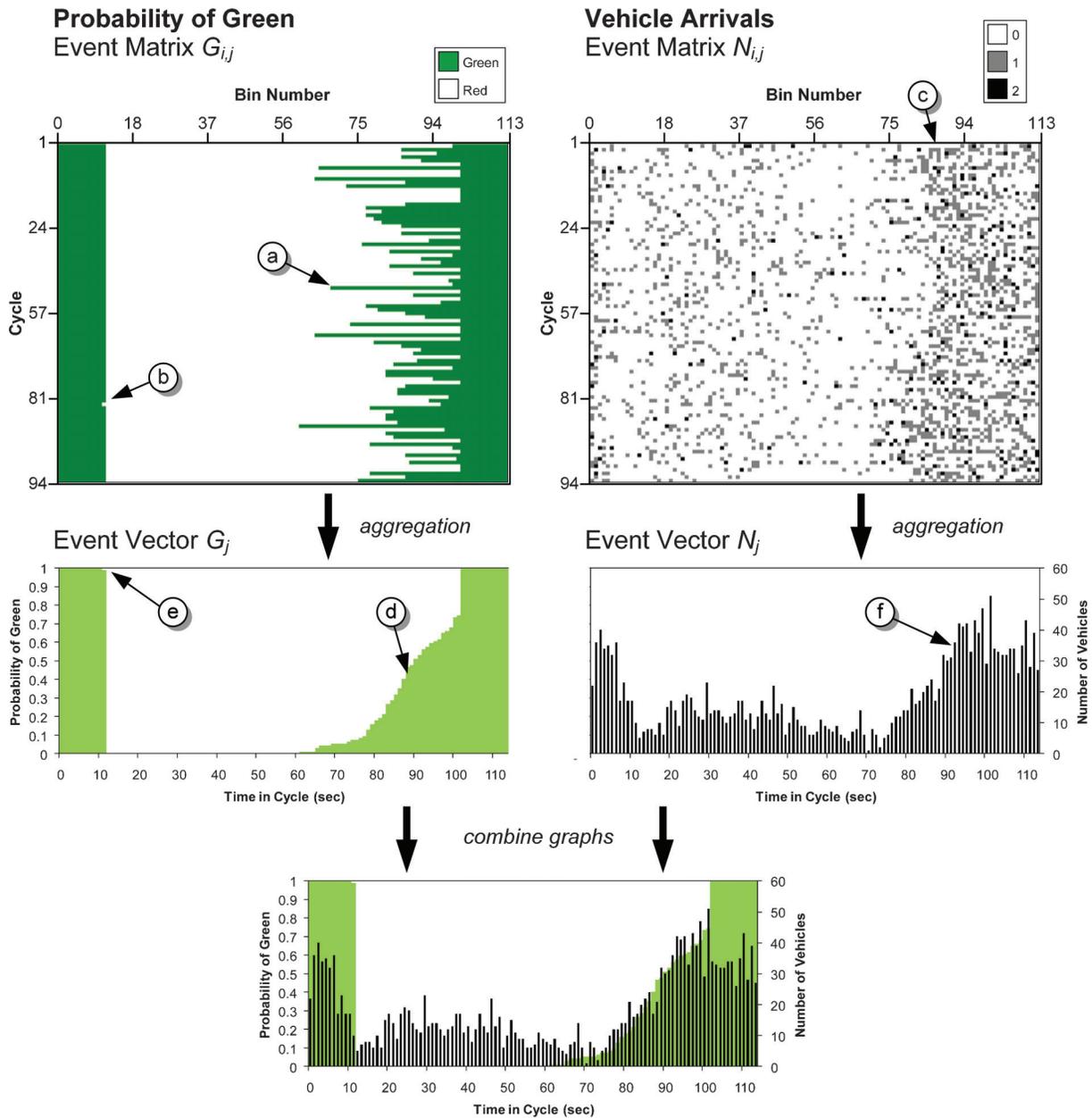


Figure 6.20 Creation of a cyclic flow profile, after Day and Bullock (60).

represents one bin in the timeline, and each cell is indexed by row i (corresponding to cycle i) and column j . Each cell has two values associated with that particular time index that store the arrival and green state: the number of vehicles arriving during the bin, $N_{i,j}$, and the status of green during that bin, $G_{i,j}$. These two matrices are illustrated at the top of Figure 6.20. These “event matrices” are transformed into “event vectors” by aggregating across each column (60):

$$N_j = \sum_i N_{i,j}, \quad (6.12)$$

$$G_j = \frac{1}{Z} \sum_i G_{i,j}. \quad (6.13)$$

This procedure is similar to the calculation of flow profiles in adaptive systems such as SCOOT and ACS-Lite (46,61,62). The quantity G_j is the probability that the signal is green during a given time in cycle. This is shown by the two distributions in Figure 6.20. The next step is to overlay those in a combined graph. This presents the status of an “average cycle” describing operations for this signal approach.

A few artifacts in the distributions are worth pointing out. Variation in the beginning of green can be seen for the individual cycles in $G_{i,j}$ (callout “a”) and in the vertical rise in the probability of green (callout “d”). There is one cycle in the analysis period when the end of green came early, as shown by callouts “b” for the individual cycle and “e” for the aggregated distribution. The block of time during which $G_j = 1$ (i.e., when the aggregated profile is at its maximum value) roughly corresponds to the green resulting from the programmed split; this is approximately 25 seconds. The start of green comes earlier about 75% of the time. In this particular example, the green wraps around the system cycle time. These observations are similar to

what can be found by graphing distributions of the beginning or end of green times (63).

In the vehicle arrival data, the coordinated platoon is visible both in the disaggregate data as a denser cloud of dots (callout “c”) and as a peak in the vehicle arrival distribution (callout “f”). The combined graph at the bottom of Figure 6.20 shows that, for this example, the platoon arrival time coincides roughly with the distribution of green. The main platoon drops off right before the end of green. The front of the platoon appears to arrive when the probability of green is less than 100%. During some cycles, the signal is still red, although there is a fair amount of early return that would, in this case, benefit the platoon at this particular intersection.

To give context to some example flow profile views, Figure 6.21 presents a PCD showing operation from 6:00 to 19:00 for phase 2 at US 36 and Post Road, for October 17, 2012. An example flow profile for the hour 7:00–8:00 is shown in Figure 6.22. It is apparent in this view that about half of the vehicle distribution is coincident with the green period. There are no clear secondary platoons in this distribution.

Figure 6.23 shows a series of flow profiles from the same data set, which are each labeled according to the one-hour spans that they represent. There are three sets of similar profiles that correspond to three different signal patterns (6:00–9:00, 9:00–15:00, and 15:00–18:00). Although the cycle length stays the same, the probability of green distribution moves around as the offsets are quite different from one plan to another. Several observations can be made from these graphics.

- In the a.m. period (6:00–9:00), the vehicle arrivals are most intense between 7:00 and 8:00. There is a considerable amount of early return to green before 7:00.
- The midday hours (9:00–15:00) are extremely consistent, and the green distribution and the arrival distribution are highly aligned, suggesting a very low percentage on

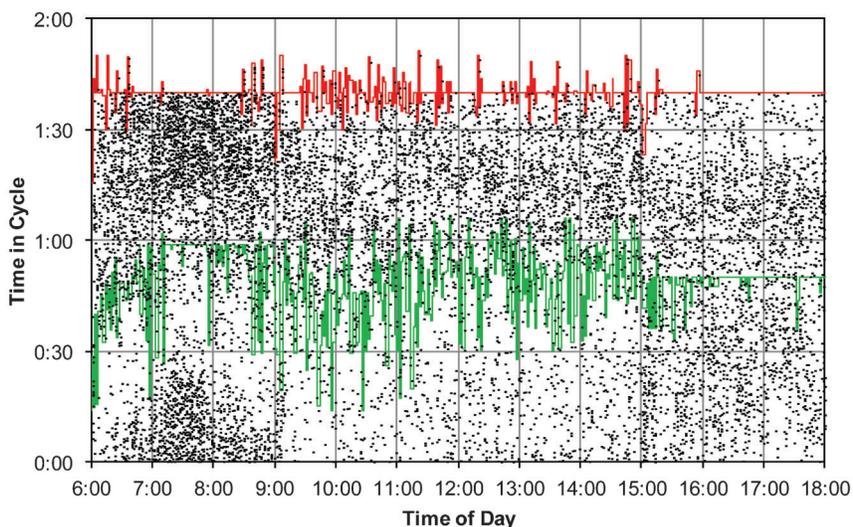


Figure 6.21 PCD with detail on the 06:00–18:00 time period.

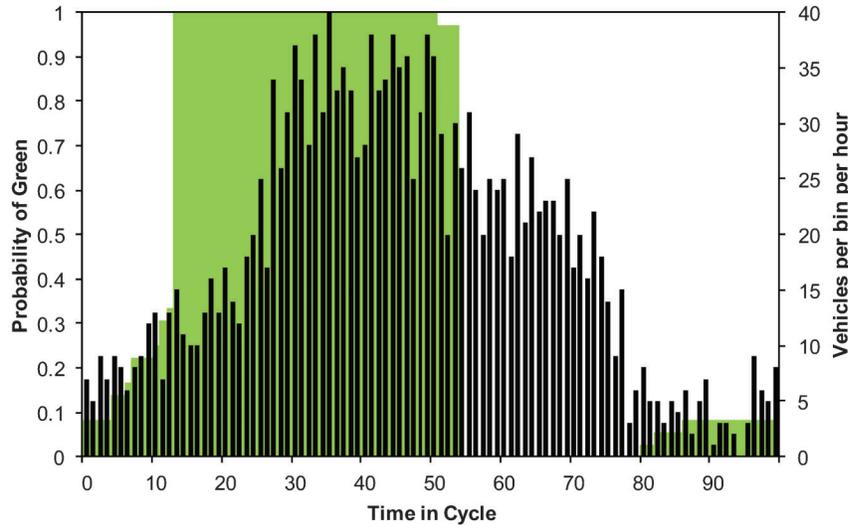


Figure 6.22 Flow profile (07:00–08:00).

green. This agrees with the percent on green that we calculated earlier (Figure 6.5).

- In the evening (15:00–19:00), there is a rather weak volume for this approach and not very well-formed platoons. It is almost impossible to perceive any substantial clusters of arrivals emerging from these distributions.

6.7 Characterizing Platoon Formation and Dispersion

If high-resolution data can be gathered at two or more adjacent intersections along a roadway, it becomes possible to observe changes to traffic flow as it progresses along the roadway from one intersection to another. Figure 6.24 shows a conceptual illustration of vehicle movement along a roadway. At the upstream intersection, platoons are formed by the release of vehicles during green; these platoons have a tendency to disperse somewhat along the roadway, owing to the variation in speed among the individual vehicles. Using the beginning and ending of green as reference points, a “search window” can be constructed where the vehicles released from the upstream can potentially be identified at a downstream detector, based on an assumed typical speed s_{typ} and a speed range based on maximum and minimum speeds (s_{max} , s_{min}). This analysis requires that the clocks of the neighboring intersections’ controllers be well synchronized.

Figure 6.25 shows an example of platoons measured using this technique along SR 37 in Noblesville, Indiana. These results are particularly interesting because they were obtained during fully actuated, *noncoordinated* operation during the late evening (22:00–24:00), yet provided some evidence for well-formed platoons during that time period (64). The percentages in the figure represent the proportion of vehicles that occur within the first 30 seconds of the distribution.

These measured distributions can be used to calibrate models of traffic flow. Figure 6.26 shows measured platoons from the same corridor, for midday operation.

The distributions (gray bars) are more spread out because the operation occurred during coordinated operations, in which the green times along the corridor are longer than during the 22:00–24:00 operation illustrated in Figure 6.25. The black lines in Figure 6.26 represent the best-fit parameters of the Robertson platoon dispersion model, based on an assumption of saturation flow at the upstream intersection. The theoretical and field-measured distributions are statistically compared to determine whether to accept or reject the model results. Collection of data across a long time period enables the model parameters to be selected for each link (65). Such information would be useful for calibrating traffic flow models, potentially improving signal timing optimization and capacity analysis (66).

6.8 Maximum Queue Length from Shockwave Estimation

This method of determining queue length was developed at the University of Minnesota (67). Under higher-volume conditions, the queue during a red phase may extend beyond the advance detectors of an intersection. In this case, using the input-output queue estimation method is inadequate because the maximum back of queue surpasses the distance to the advance detector from the stop bar. To calculate the maximum back of queue, the speeds of two shockwaves (arrival and departure) determine when the queues are the longest and how far beyond the advance detector the queue has reached. The speed of each shockwave is computed using the phase begin time and detector presence times logged by the signal controller.

In Figure 6.27, after the phase turns red, a queue forms as vehicles arrive at the intersection. The arriving vehicles produce a queuing shockwave that propagates upstream along the link at velocity V_1 . In Figure 6.28, the queuing shockwave has reached past the advance detectors at 405 ft and continues further upstream. The

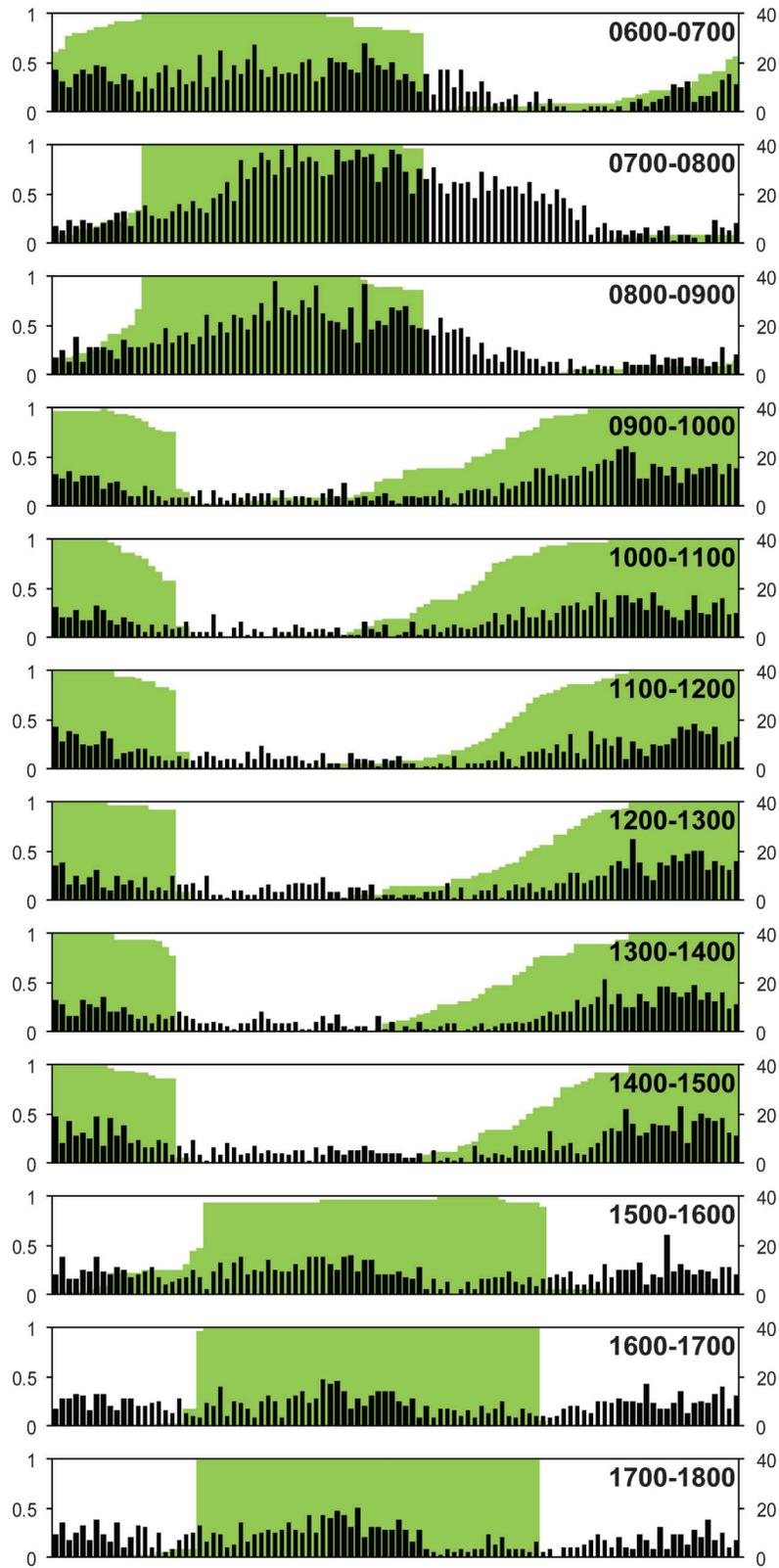


Figure 6.23 Flow profiles by hour.

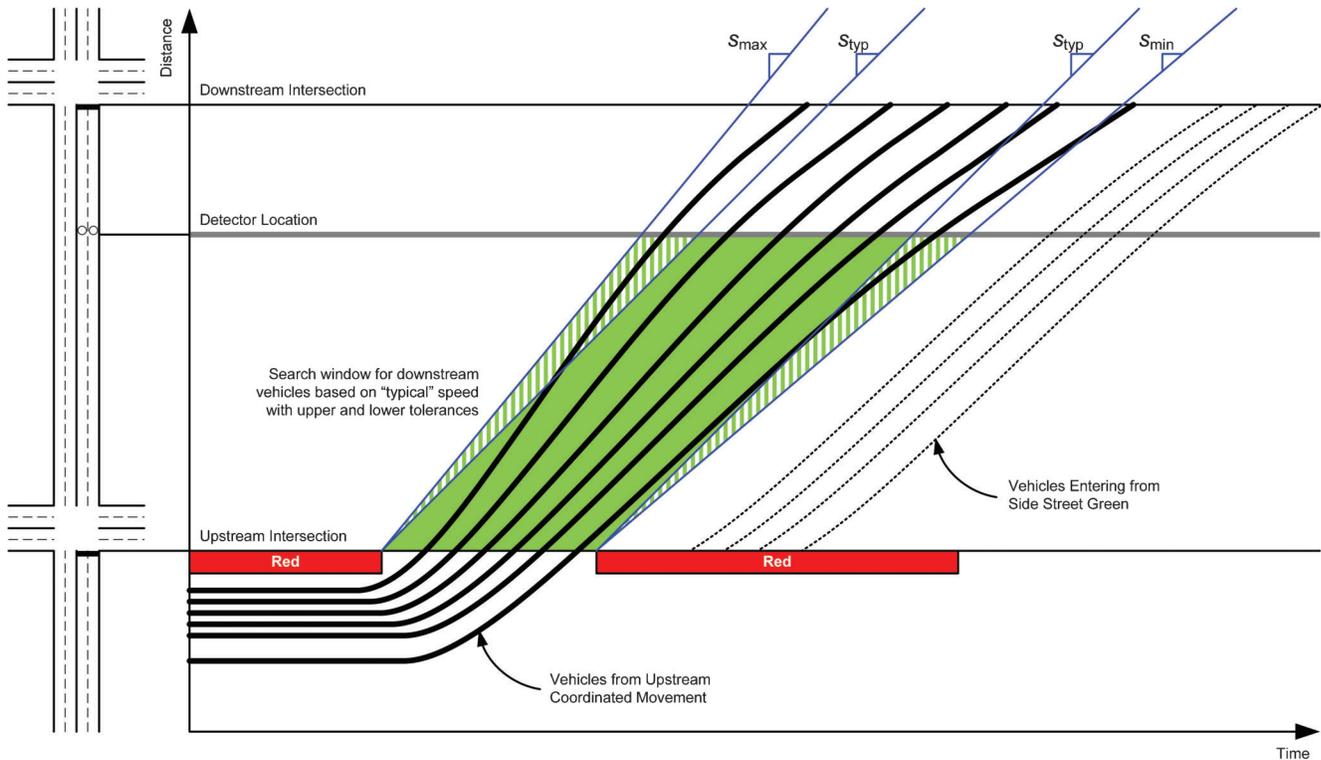


Figure 6.24 Finding downstream vehicles that are likely to have departed from the upstream intersection through the coordinated movement (65).

stationary vehicles at the advance detectors trigger “detector-on” events during the formation of the queue, but do not trigger “detector-off” events until after the vehicles start moving.

After the phase turns green, vehicles depart the intersection at the saturation flow rate. The departures produce a second shockwave propagating up the link with velocity V_2 . As the departure shockwave passes the advance detectors, a “detector-off” event is triggered. The speed of the departing shockwave V_2 can be computed using the distance from the stop bar to the advance detectors over the time interval of phase green begin to the detector-off event. As the departure shockwave continues down the link, at some point it will reach the last stationary vehicle where it meets the arrival (queuing) shockwave. The arrival flow rate is assumed to be lower than the departure (saturation) flow rate for this to be possible. At this point, the queue has reached the maximum distance from the stop bar.

In Figure 6.29, a third shockwave with velocity V_3 propagates toward the stop bar. This shockwave is the moving “flow boundary” between the departure and arrival flow rates. As vehicles previously stopped in the queue move toward the stop bar at saturation flow, the vehicles behind them, having not been queued previously, travel at a lower rate, which is the arrival flow. As V_3 propagates toward the stop bar, the change in flow volume can be measured by the advance detectors using changes in headway durations. Once the headways increase beyond a predefined threshold, this point

in time can be marked as the arrival of the third shockwave from the back of the queue. The third shockwave’s speed V_3 and distance to the maximum back of queue can be estimated based on the speed and arrival of the departure shockwave and the arrival flow rate computed after the “flow boundary” shockwave has passed. Figure 6.30 illustrates the continued propagation of V_3 toward the stop bar.

Table 6.4 presents 32 high-resolution events during high-volume conditions where the queue length has surpassed the distance to the advance detector from the stop bar. The vehicle arriving at the first event triggers the detector at 7:26.17.4 and occupies the detector for 16.3 seconds until the off event at 7:26:33.7. The long occupation interval qualifies as an indication of a long queue. Depending on how the metric is implemented, the threshold for the occupation interval can vary.

As the first vehicle departs from the detector, the departure shockwave continues down the link at velocity V_2 . The headways of subsequent vehicles are recorded and averaged as the departure flow rate variable. Once a vehicle’s headway is detected to be significantly greater than the departure average, the time of that discrepancy is considered to be the time that the “flow boundary” shockwave propagates to the advance detector from the back of queue. This is indicated by event 18 in Table 6.4, where the headway of the arriving vehicle jumps to 9.3 seconds, significantly larger than the previous 2.3-second average. From this point on, the headways of all vehicles

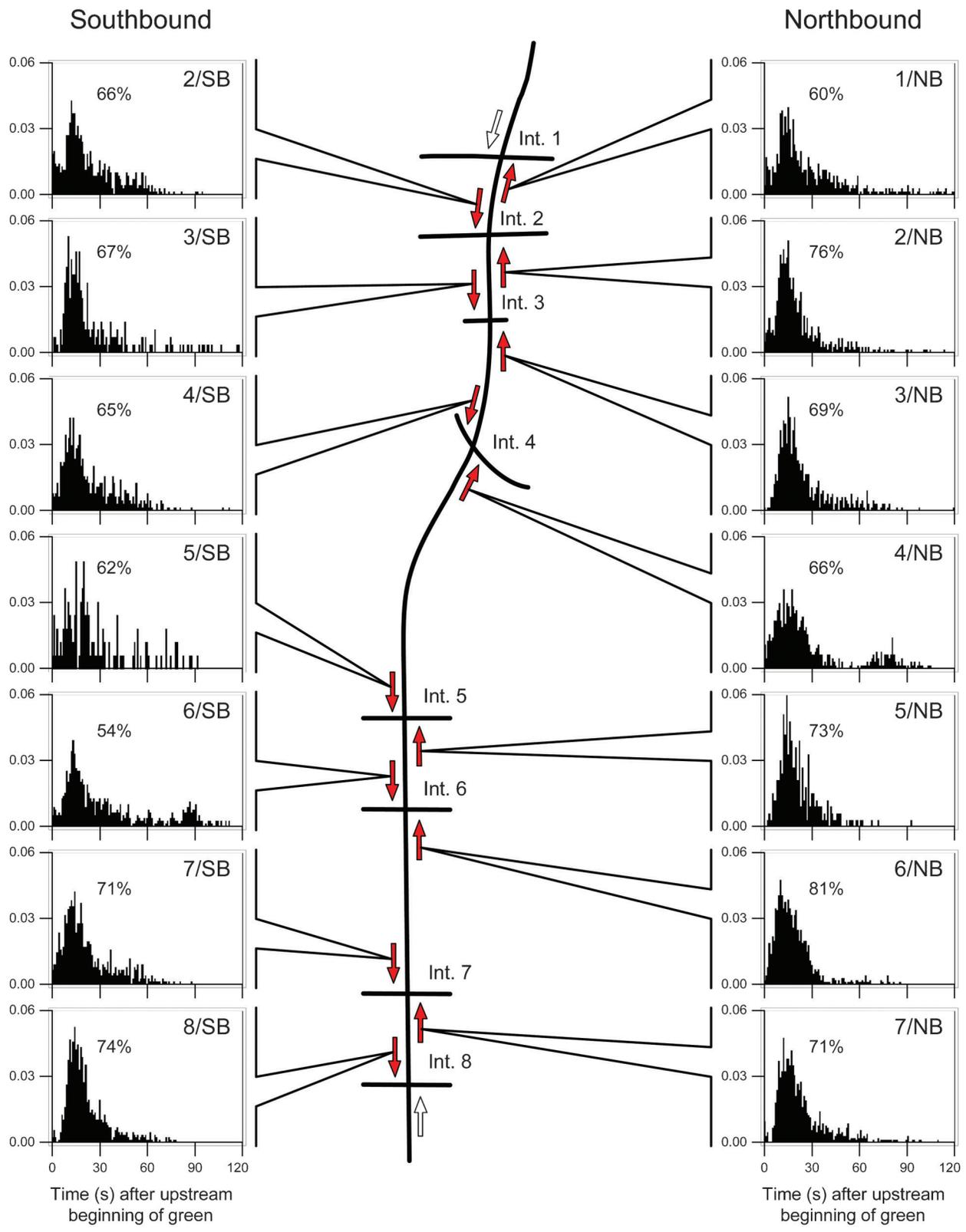


Figure 6.25 Measured platoons based on upstream beginning of green during fully actuated, noncoordinated operation: SR 37 in Noblesville, Indiana, June 30, 2010, 22:00–24:00 (64).

$\checkmark+$ accept at 99% $\checkmark-$ accept at 90% \times reject at 90%

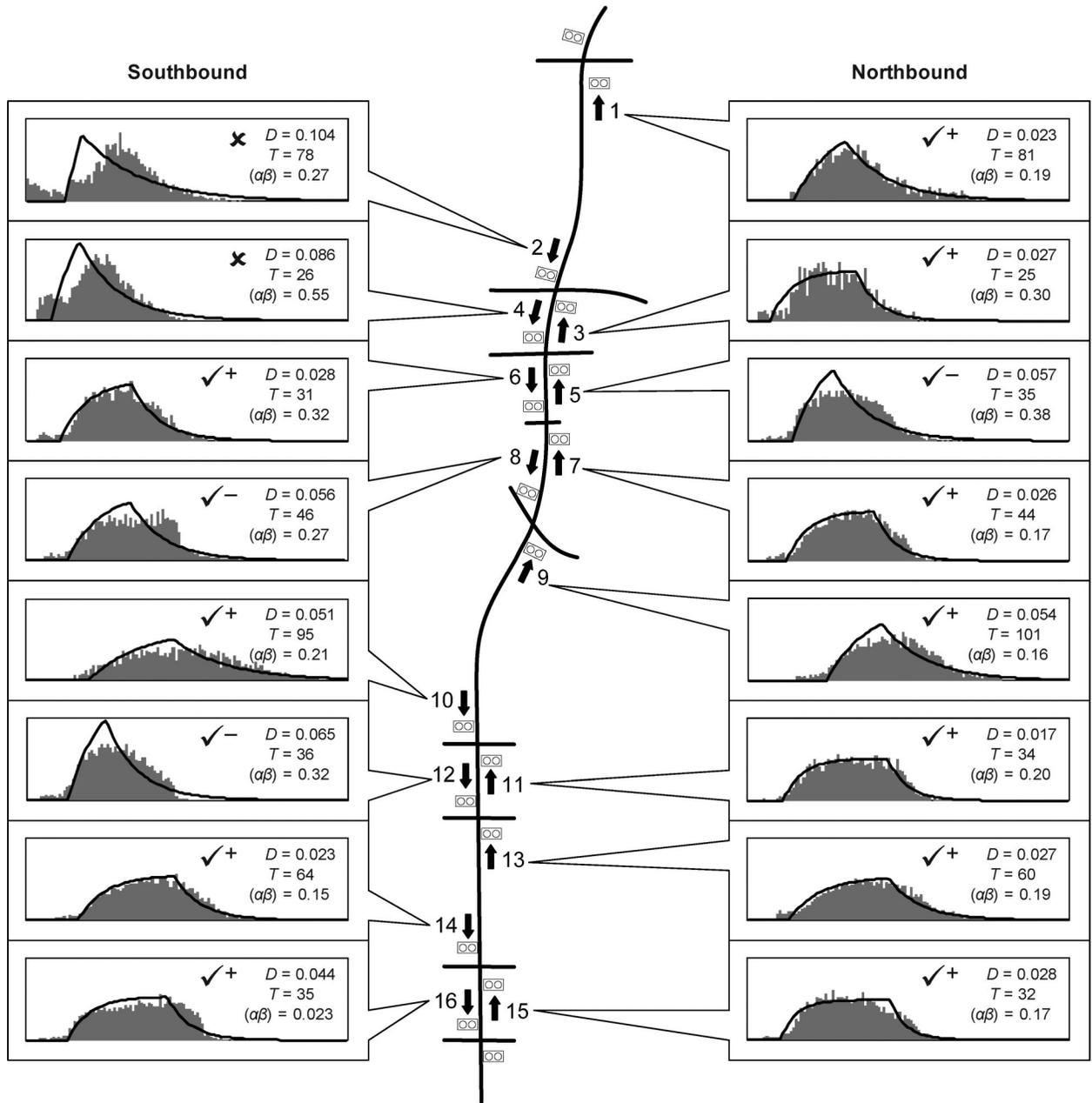


Figure 6.26 Matching measured platoons against modeled platoons: example results from SR 37, for data from March 12, 2011, 14:00–16:00 (65).

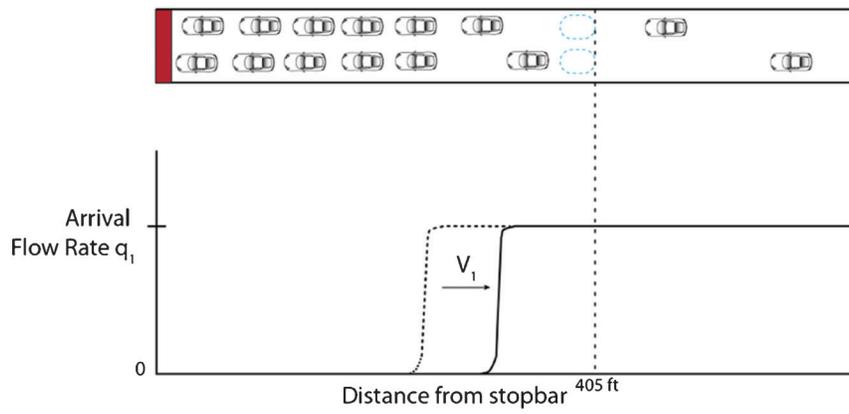


Figure 6.27 Queuing shockwave approaching advance detector.

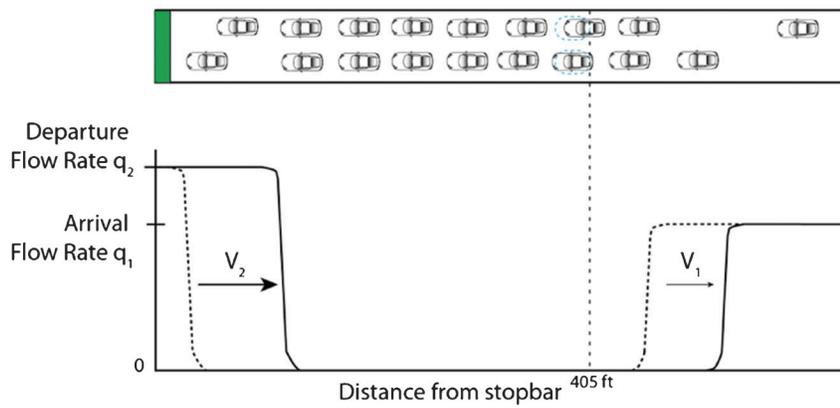


Figure 6.28 Queue occupying advance detector as phase turns green and more vehicles arrive.

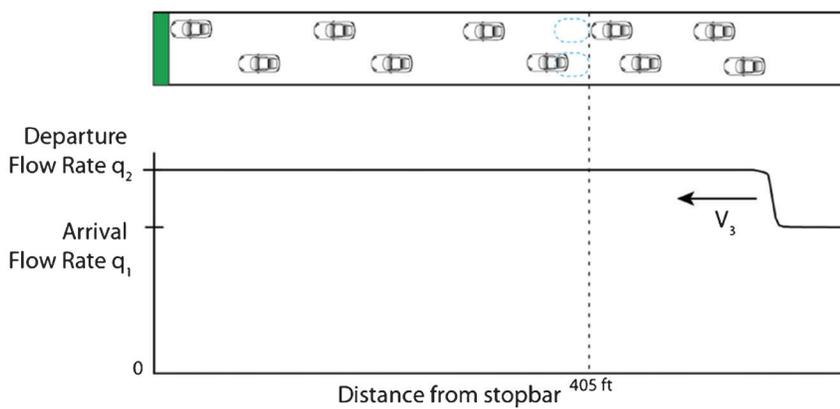


Figure 6.29 Saturated flow departures reaching the maximum back of queue.

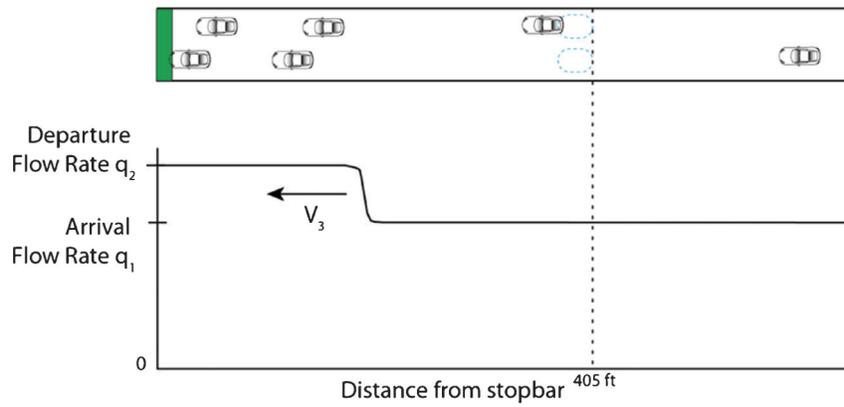


Figure 6.30 Flow rate change once queue is discharged past advance detector.

TABLE 6.4.
Using high-resolution event data to determine departure and arrival shockwaves.

k	Timestamp	Event	“On” Time Duration	h	Departure h Average	Arrival h Average
1	7:26:17.4	Detector On				
2	7:26:19.0	Begin Green				
3	7:26:33.7	Detector Off	16.3			
4	7:26:35.1	Detector On				
5	7:26:36.3	Detector Off	1.2			
6	7:26:37.4	Detector On		2.3	2.3	0.0
7	7:26:38.4	Detector Off	1.0			
8	7:26:39.8	Detector On		2.4	2.4	0.0
9	7:26:40.6	Detector Off	0.8			
10	7:26:41.7	Detector On		1.9	2.1	0.0
11	7:26:42.6	Detector Off	0.9			
12	7:26:44.1	Detector On		2.4	2.3	0.0
13	7:26:44.8	Detector Off	0.7			
14	7:26:46.6	Detector On		2.5	2.4	0.0
15	7:26:47.2	Detector Off	0.6			
16	7:26:48.8	Detector On		2.2	2.3	0.0
17	7:26:49.4	Detector Off	0.6			
18	7:26:58.1	Detector On		9.3	2.3	9.3
19	7:26:58.5	Detector Off	0.4			
20	7:27:01.2	Detector On		3.1	2.3	6.2
21	7:27:01.5	Detector Off	0.3			
22	7:27:03.5	Detector On		2.3	2.3	4.2
23	7:27:03.9	Detector Off	0.4			
24	7:27:05.4	Detector On		1.9	2.3	3.1
25	7:27:05.9	Detector Off	0.5			
26	7:27:09.3	Detector On		3.9	2.3	3.5
27	7:27:09.7	Detector Off	0.4			
28	7:27:13.1	Detector On		3.8	2.3	3.6
29	7:27:13.6	Detector Off	0.5			
30	7:27:17.8	Detector On		4.7	2.3	4.2
31	7:27:18.2	Detector Off	0.4			
32	7:27:22.4	Detector On		4.6	2.3	4.4

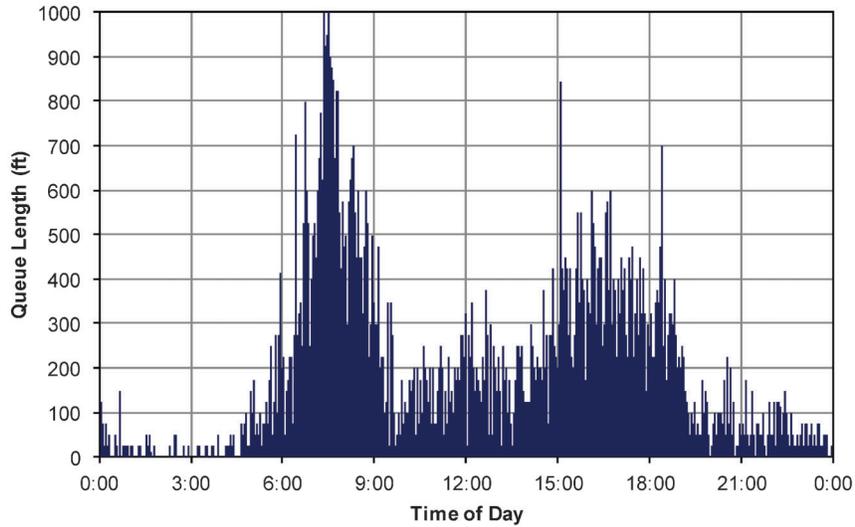


Figure 6.31 Estimated queue lengths.

arriving are considered to be the arrival headways and are averaged under the arrival headway variable. The number of headway samples for computing the average arrival flow rate can be set as a fixed threshold, or as a function of the difference from the departure flow rate. If the average arrival headway differs by more than a predefined threshold from the departure flow rate, a cutoff point can be triggered to finalize the calculations. After the arrival and departure flows have been calculated, the maximum back of queue can be calculated from

$$L_{\max} = L_{\det} + \frac{T_c - T_b}{\left(\frac{1}{V_2} + \frac{1}{V_3}\right)}, \quad (6.14)$$

where L_{\max} is the maximum length of the queue, L_{\det} is the distance between the advance detector and the stop bar, T_c is the time the flow boundary shockwave arrives at the advance detector, T_b is the time the departure shockwave arrives at the advance detector, V_2 is the velocity of the departure shockwave, and V_3 is the velocity of the flow boundary shockwave. Figure 6.31 illustrates queue lengths computed throughout one day using the described metric for westbound phase 2 at US 36 and Post Road on October 17, 2012. It is helpful to compare this against the plot of average delay in Figure 6.11. Both plots show substantial queues or delay during the a.m. peak, especially around 7:00–8:00, as well as a modest increase during the afternoon.

7. MULTIMODAL PERFORMANCE MEASURES

This chapter presents performance measures for nonvehicle modes at traffic signal systems. The three modes investigated here are pedestrians, railroad, and transit. These modes are accommodated in signal controllers by pedestrian phasing, railroad preemption, and transit priority. These aspects of controller operations have detection capabilities and performance

considerations that can be quite different from motor vehicles.

7.1 Pedestrian Demand

High-resolution data have been investigated in the context of pedestrian operations in a few previous studies (68,69). At most intersections featuring pedestrian phases, there is usually no system for pedestrian detection other than pedestrian pushbuttons. This situation may change as new detection technologies become available (and affordable), but it is unlikely that much will change in the foreseeable future. Other intersections do not feature pedestrian pushbuttons or indications. However, it is still possible in this case to at least characterize the potential level of service for pedestrians by measuring the conflicting vehicular volume.

Except for intersections with an exclusive pedestrian phase, pedestrian movements are usually concurrent with an adjacent through movement. Two vehicular movements usually conflict with the pedestrian movement: right turns from the same approach as the adjacent through movement, and opposing left turns that occur during the permitted phase. Although technically these vehicles are supposed to yield to pedestrians, incursions are not uncommon when traffic levels are high and/or the level of visibility is poor. The problem is a bit more severe for the permitted left turns, which must yield both to pedestrians and to opposing vehicular traffic. A vehicle that has accepted a gap in vehicular traffic may face a dilemma if the driver has misjudged a gap in the pedestrian traffic, either forcing an incursion into the pedestrian right-of-way or forcing the opposing traffic to stop. The flow rates of conflicting traffic characterize the overall pressure on the pedestrian movements.

The data presented here were collected in September 2007 at the intersection of Northwestern Avenue and

Stadium Drive in West Lafayette, Indiana. These data were selected because, at the time of writing, they were the only available field data with a significant amount of pedestrian traffic for analysis. The intersection phasing was changed to feature an exclusive pedestrian phase in 2009, and it was desired to investigate performance under the more typical preexisting state. Figure 7.1 shows an overview of this intersection, including the locations of vehicle detectors.

Figure 7.2 shows a measurement of the percentage of cycles in which a pedestrian phase was served, for the four pedestrian phases at Northwestern and Stadium. The data are sorted into 30-minute bins. As expected, there are very few actuations of the pedestrian phases during the early morning; there is some activity throughout most of the day, but the peak tends to occur around the middle of the day, which is the time when there is substantial activity in the surrounding area. This intersection is located near a college campus, and pedestrian activity is sustained relatively late into the evening.

If the phase actuation times (beginning of the active pedestrian phase call) are compared with the beginning of the walk time, an estimation of the pedestrian delay can be obtained (68). Here, this is called the “service-to-actuation” time. Figure 7.3 shows a view of these data for eastbound phase 4 at Northwestern and Stadium. There is a considerable amount of variation in this value; throughout the day, the values range from as low as a few seconds to as long as 126 seconds. The longer delays are likely occurrences where a pedestrian pushed the button after the beginning of the “flashing don’t walk” phase. It is not possible to discern from these data whether the pedestrian actually waited until the next walk phase to actually cross the street, but these data represents the

time to when he or she would have received the right-of-way. The range of values is proportional to the cycle length at the intersection.

Figure 7.4 and Figure 7.5 are alternate views of the same baseline data, which show the number of vehicles crossing the eastbound pedestrian movement. Figure 7.4 expresses the conflicting volume in terms of vehicles per hour, and Figure 7.5 has converted this into the equivalent average vehicle headway in terms of seconds per vehicle. Figure 7.5 shows the average gap available to pedestrians attempting to execute this movement. During many cycles, this becomes as low as 2 seconds, which is representative of sustained saturation flow during green. This occurs at the same time as the peak in the pedestrian demand for that movement. From this, it is possible to conclude that there is considerable pressure on the pedestrian movement at this intersection. This type of data could identify locations where pedestrian movements might potentially benefit from treatments such as a leading pedestrian interval or an exclusive pedestrian phase.

7.2 Preemption and Priority

At some locations, traffic signals receive requests for preemption or priority from railroads, emergency vehicles, or transit vehicles. Preemption and priority are both requests for a particular intersection display. High-level preemption requests take precedence over all other signal operations, and are reserved for “emergency” operations. Generally, preemption failures are unacceptable—especially for situations such as an adjacent railroad crossing or drawbridge. Priority requests, on the other hand, are less critical. While it is desirable to achieve a certain display early (e.g., to reduce delay for a transit vehicle), a disaster will not occur should the request be unsuccessful.

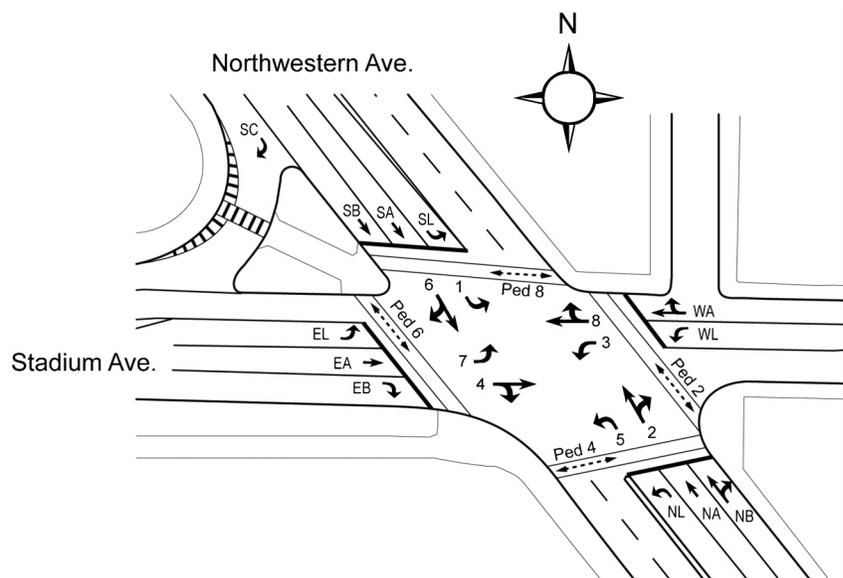
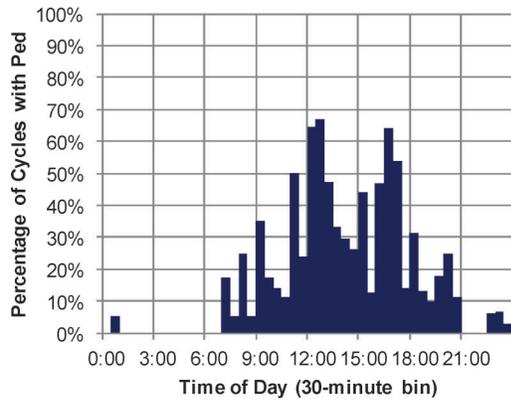
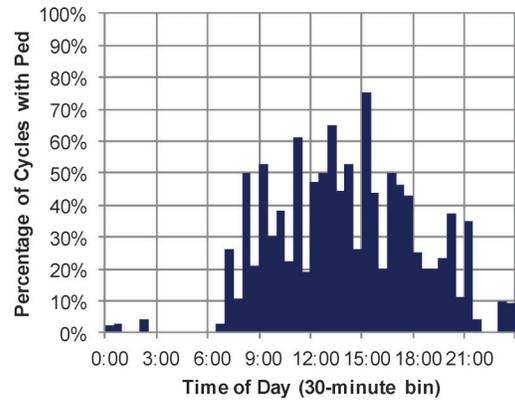


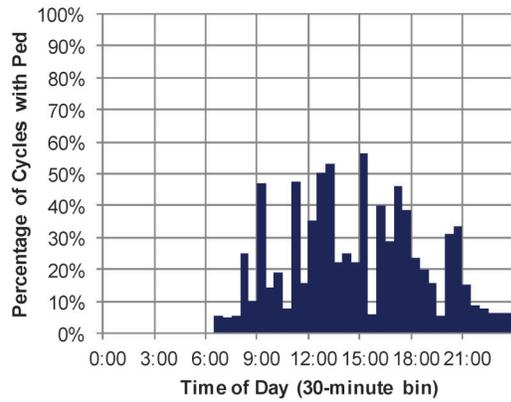
Figure 7.1 Northwestern Avenue and Stadium Drive, West Lafayette, Indiana.



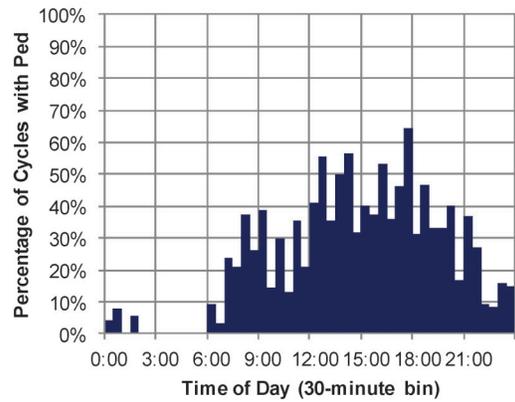
(a) Northbound phase 2



(b) Southbound phase 6



(c) Eastbound phase 4



(d) Westbound phase 8

Figure 7.2 Percentage of cycles with pedestrian phases at Northwestern and Stadium, September 19, 2007.

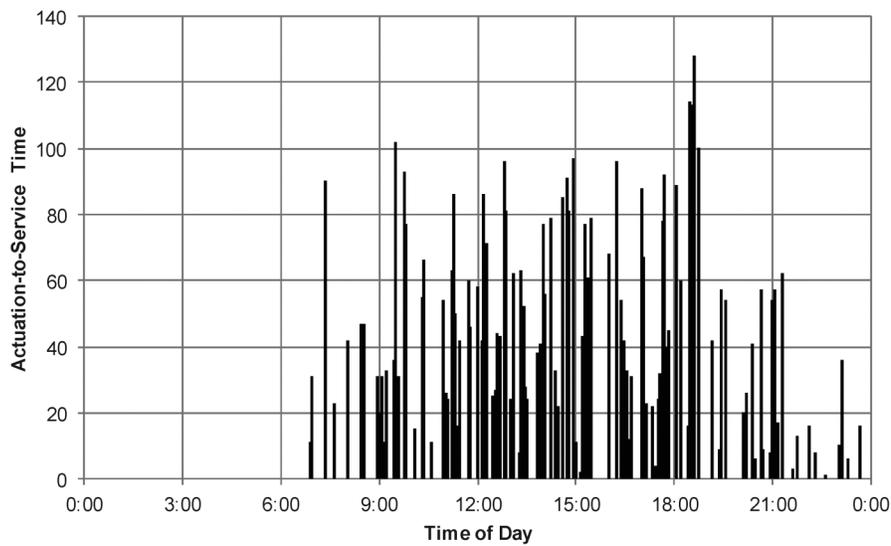


Figure 7.3 Actuation-to-service time, phase 4, Northwestern and Stadium, September 19, 2007.

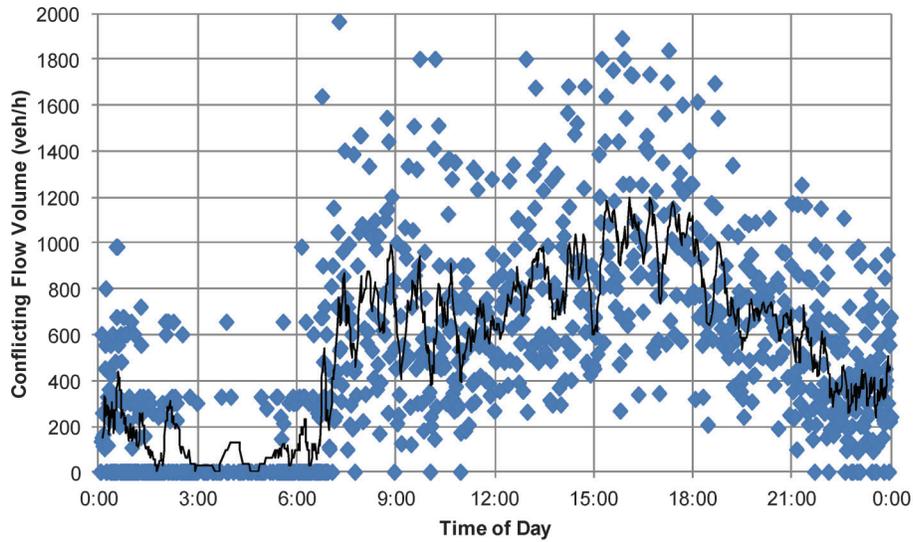


Figure 7.4 Conflicting volume, phase 4, Northwestern and Stadium, September 19, 2007.

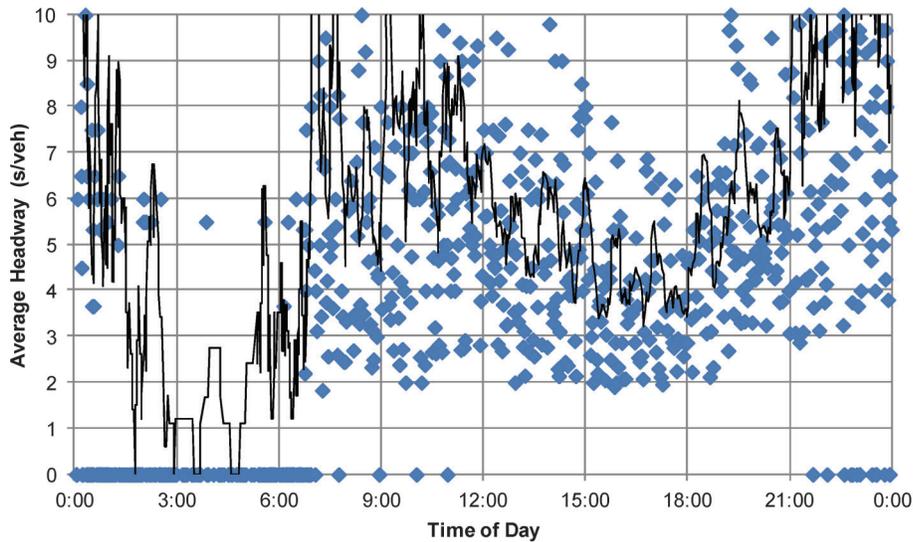


Figure 7.5 Conflicting volume, phase 4, Northwestern and Stadium, September 19, 2007.

Several performance measures for railroad preemption were discussed in previous studies (70,71). These are based on event times for various intervals defined during the preemption process (Figure 3.20). It is particularly important to verify that the gates at an adjacent railroad crossing descend before the termination of the track clearance green, to ensure that vehicles queued on the tracks are served a green indication that permits them to move forward.

Figure 7.6 shows an event diagram revealing the performance of individual preemption events throughout the day at the intersection of US 36 and Carroll Road in Indianapolis, Indiana. This intersection has an approach with a railroad crossing within 100 ft of the

intersection. Because there are relatively few preemptions occurring during the day, a view of each individual preemption event timeline can feasibly be shown in one graph. Here, the Gate Down event occurs before the termination of track clear green during every preemption event throughout the day; this verifies that preemption is operating as expected. There is adequate time after the gate descent to clear any traffic standing in front of the tracks.

The total duration of each preemption event is shown in Figure 7.7. This covers the span of time from the beginning of the preempt call until the end of limited service. Train traffic is clearly at its heaviest around 9:00 during this day. The events themselves vary

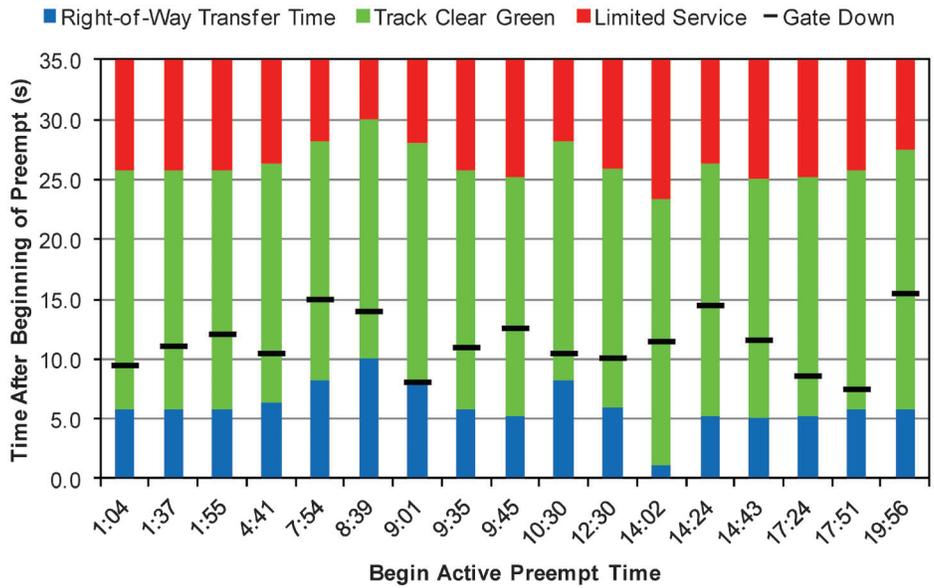


Figure 7.6 Event diagrams for preemptions at US 36 and Carroll Road.

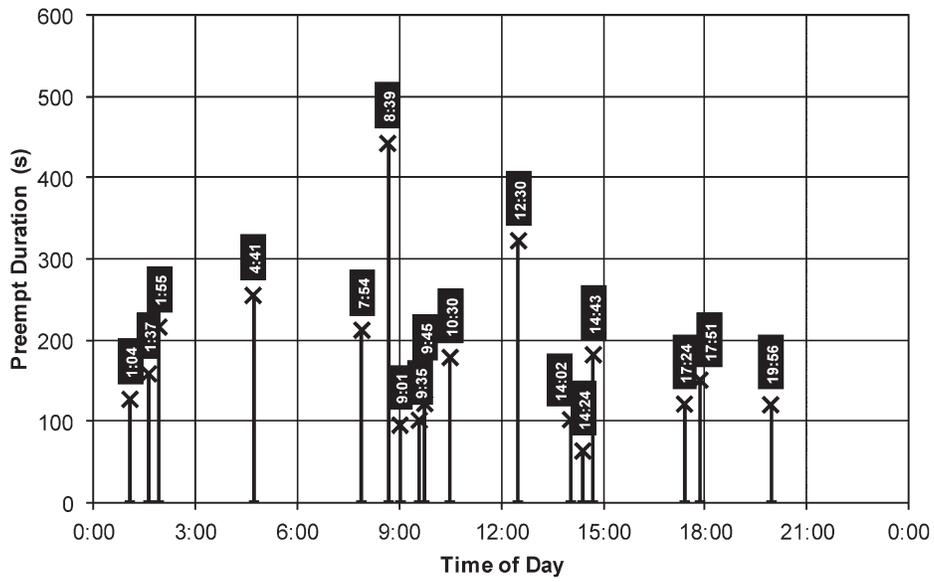


Figure 7.7 Duration of preempt events at US 36 and Carroll Road.

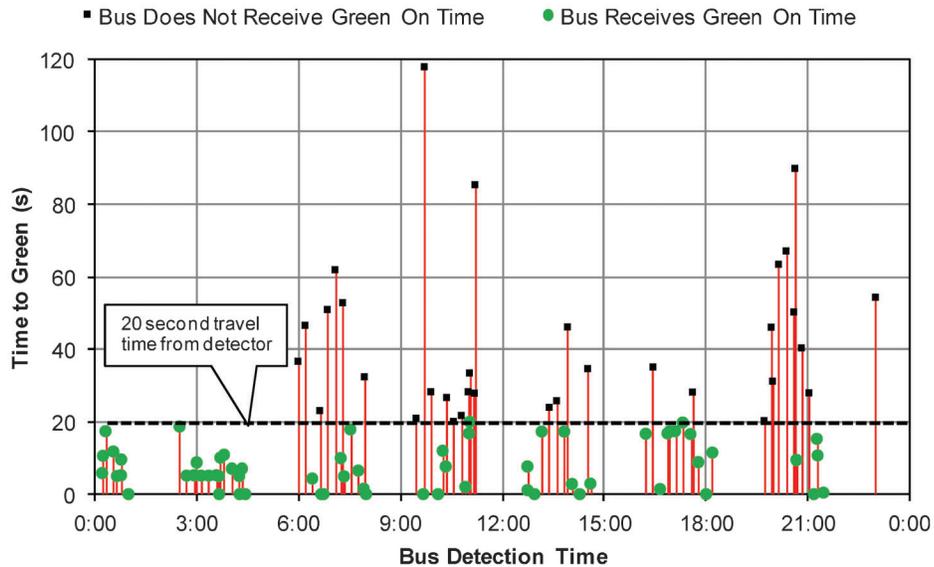


Figure 7.8 Bus performance with transit signal priority (simulation data).

in their duration from about one minute to longer than seven minutes.

Figure 7.8 shows a similar graph³ that gives the time between the inception of a priority request and the beginning of green for the desired movement. Priority requests are located 20 seconds upstream from the stop line; if the indication for the bus movement changed to green before 20 seconds, it is likely that the bus did not have to stop. For most of the day, the priority settings usually permit the bus to traverse the intersection with a low likelihood of stopping. Furthermore, during most of those instances where the bus was likely to have stopped, the controller was able to serve the priority movement within 40 seconds of the beginning of the request. There were a few instances where it took longer, particularly around 6:00 and 21:00, likely because of the bunching of requests during those times (i.e., many subsequent requests take place appear to occur around those times).

8. MAINTENANCE PERFORMANCE MEASURES

In addition to operational performance, maintenance is important to ensure that the traffic control equipment is functioning properly. In particular, it is important to maintain communication and detection equipment to support a performance measurement system. Detection is important for understanding the relative volumes for vehicles and other modes using the intersection, and communication is essential for

³At the time of writing, data were not yet available from an intersection where transit signal priority was in use. The data used to create this graph were generated in simulation.

returning these data to a traffic management center for analysis.

8.1 Communications Quality

“Communications” means the electronic linkage between the traffic signal controller and the traffic management center (see Figure 1.6). This may vary from direct fiber-optic lines to radio IP connections. Performance measures for communications quality were explored in a recent paper (45) to analyze the effectiveness of wireless IP communications throughout a state-wide network. This section is based upon that paper.

Data quality can be measured using the ping utility as well as assessing data completeness of successfully downloaded data. “Pinging” works by measuring the roundtrip time of test data packets sent out over the network to a specific destination from a particular source. In the case of a signal network, a server at a traffic management center typically initializes the ping mechanism to contact controllers out in the field. Ping success generally relies on having a working route with adequate capacity between both endpoints in the system at the time the test is conducted. Reliability of ping is affected by factors such as physical severance of the connection, routing and firewall issues, network connection problems, and power outages. While ping gives an idea of the network conditions at any given moment in time, data completeness quantifies how much data have actually been brought into the central system through the network.

Figure 8.1 plots data completeness, ping, and the expected “ideal” data trendline for a controller during a 48-hour period measured in hour increments. The Ideal line is the quantity of data or ping successes that is expected during the period—here the rate of success in

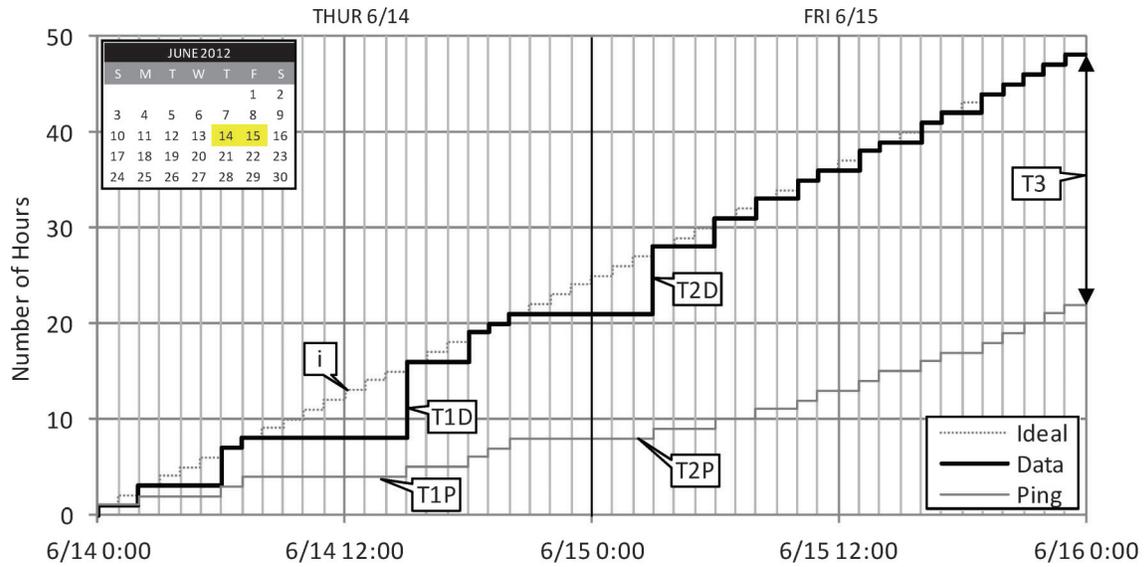


Figure 8.1 Cumulative data and communication activity at State Route 37 and Banta Rd (45).

a fully functional system should increase by one increment per hour for both ping and data. In Figure 8.1 at callout “i,” the controller lagged behind in both data and ping success for the first 12 hours: only 8 hours of data completeness and 4 hours of ping successes were recorded. At callout “T1D,” data were recovered up to the Ideal line owing to a brief recovery of communications, as indicated by the Ping line at 15:00 on June 14. This was possible because, during communications downtime, the controller was continuously buffering event data. When the connection was reestablished, data that have not been uploaded were retrieved from the buffer. This process occurred again at Callout T2D. By the end of the two days, the data were fully recovered to its ideal state, but in terms of ping successes during that period, there had only been 22 hours of working communications. The gap between

data downloaded and the ping uptime demonstrates the resilience of signal networks against communication failures.

8.2 Data Completeness

In a central management database, both temporal and spatial criteria must be established to determine whether or not a set of data satisfies a specific completeness threshold. Figure 8.2 illustrates data completeness in the context of a two-and-a-half month period for 123 intersections with communication connectivity for the State of Indiana. The green bars represent the number of intersections with data recorded for each day. The blue bars represent the number of intersections with 24-hour communication for each day. Local data buffering within a high-latency network experiencing occasional service

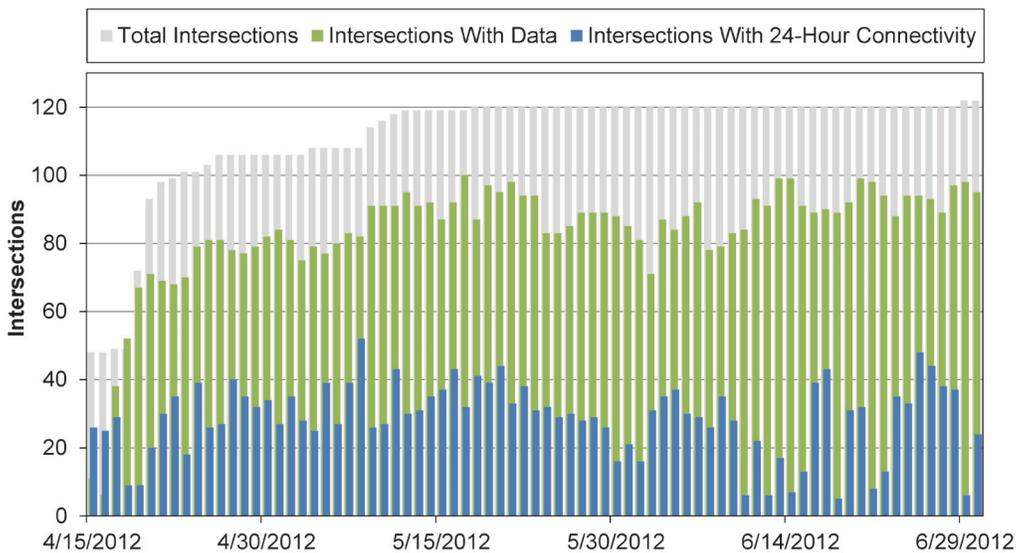


Figure 8.2 Longitudinal analysis of infrastructure and application data for entire system (45).

disruptions have demonstrated the resilience of the data upload process. The chart also reveals that there are about 20 intersections defined in the monitoring system but not yet fully integrated because of issues such as local communications equipment failure or pending controller firmware upgrades.

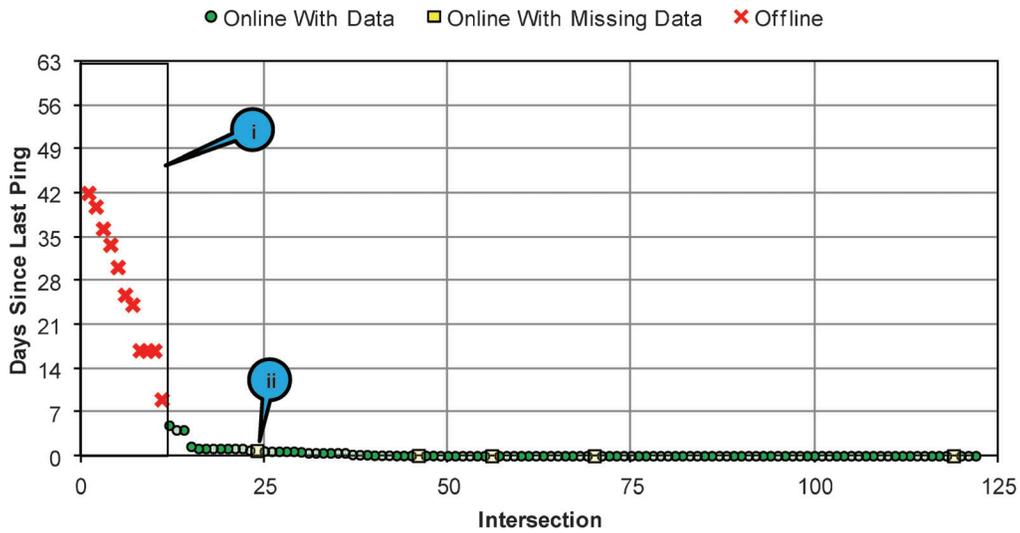
A controller that is set up to log and upload data to a central management system can exhibit four different states:

1. A controller where both ping and data collection has been successful is online.
2. A controller that responds to ping tests but has data missing is likely to be experiencing a configuration error (e.g., data logging is disabled or the FTP application is unresponsive).
3. If the ping test is unsuccessful but data are present, it is possible that data have been manually uploaded to the

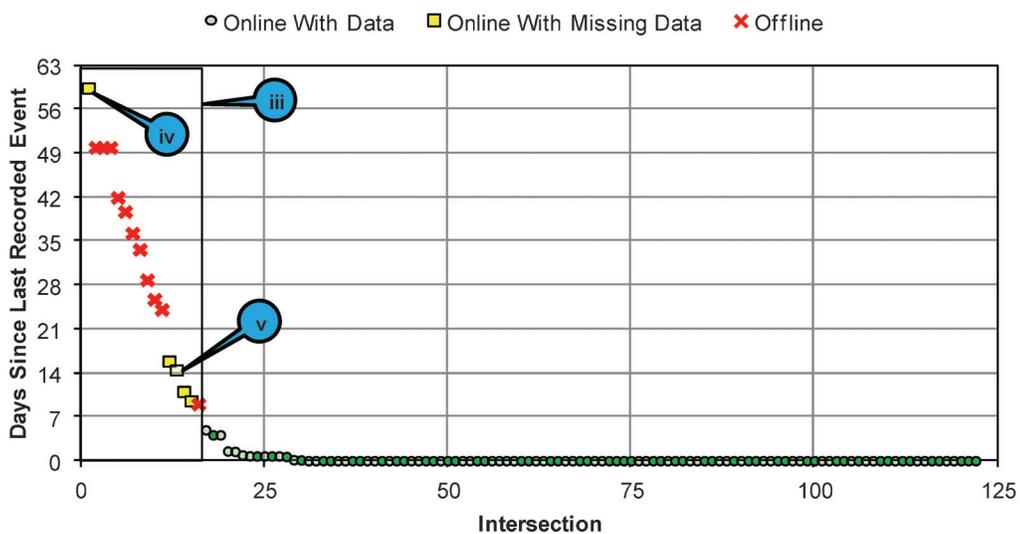
server (e.g., by transferring on a flash drive), or otherwise the ping application has stopped working.

4. An intersection that both fails the ping test and has no data is offline.

Figure 8.3 provides an overview of communication performance in the system as of July 1, 2012. The Pareto diagrams show the proportion of intersections in the system malfunctioning relative to the rest of the system. Figure 8.3a ranks intersections by the number of days since the most recent ping. Intersections with no successful ping for more than seven days are considered offline (Figure 8.3a, callout “i”). The remaining intersections are all online, but five are missing data (Figure 8.3a, callout “ii”). Figure 8.3b shows the same data ranked by the number of days since the last recorded data event, which reflects the most recent successful FTP data transmission.



(a) Number of days since previous successful ping



(b) Number of days since previous successful data insertion

Figure 8.3 Determining deficiencies in current communications health (45).

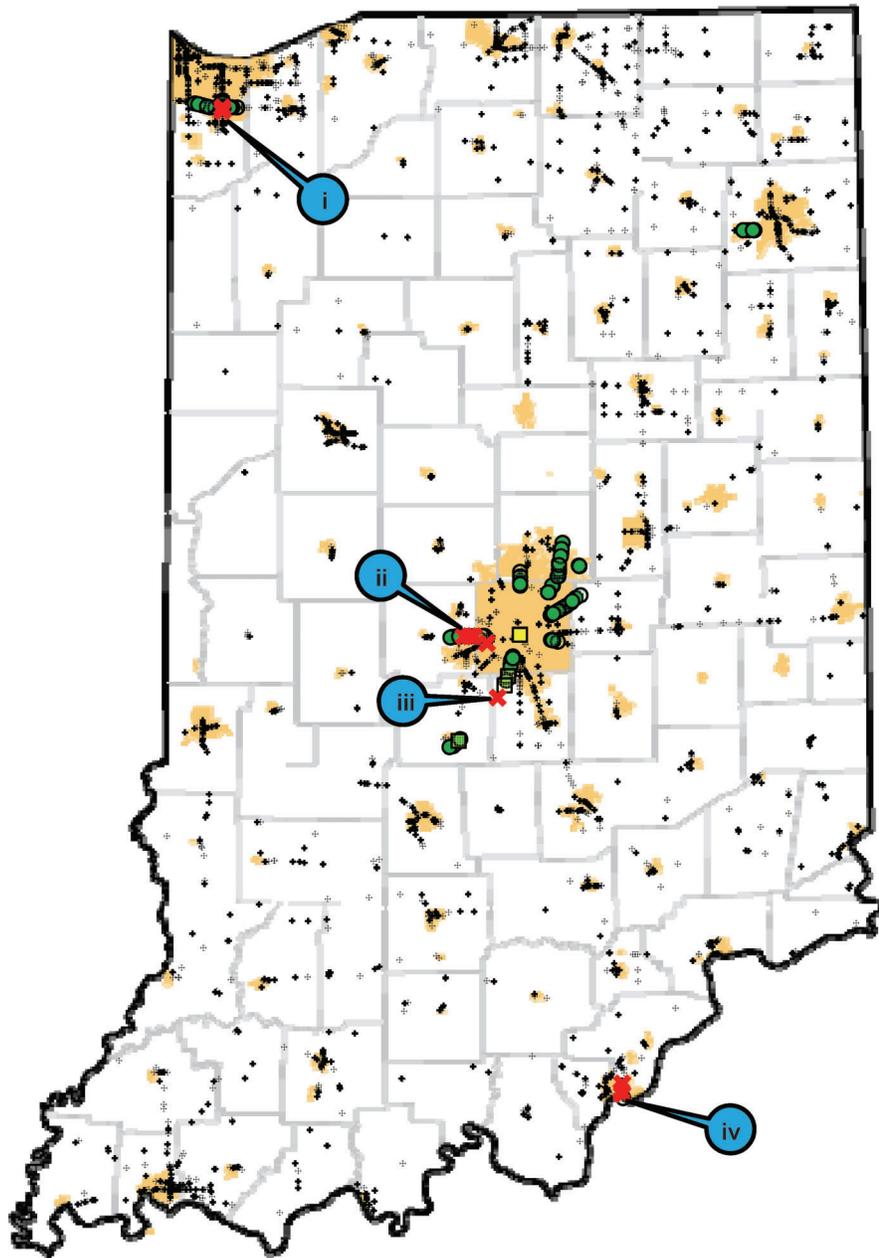


Figure 8.4 Mapping deficiencies in the data collection infrastructure (45).

Most of the intersections missing data for seven or more days are offline (Figure 8.3b, callout “iii”). However, a few of these are still online but missing data, indicating a layer 7 failure (Figure 8.3b, callouts “iv” and “v”).

Figure 8.4 shows system status in a geographic map of the INDOT signal network. The intersections with communication problems are located in four specific corridors as indicated by the callouts. It is not difficult to visualize many of the remaining currently excluded intersections also reporting status. This type of visualization would help agencies assess qualitatively the need for maintenance and network improvements throughout the system.

8.3 Detector Failures

It is impossible for traffic control systems to operate efficiently without working subsystems, especially detector systems. This is especially true for more sophisticated control systems. Performance measures for monitoring detector status are therefore a vital component of an overall portfolio of maintenance performance measures. Signal controllers and detection systems typically feature some functions for reporting a variety of errors. However, usually this information is not easy to acquire remotely and is not always logged

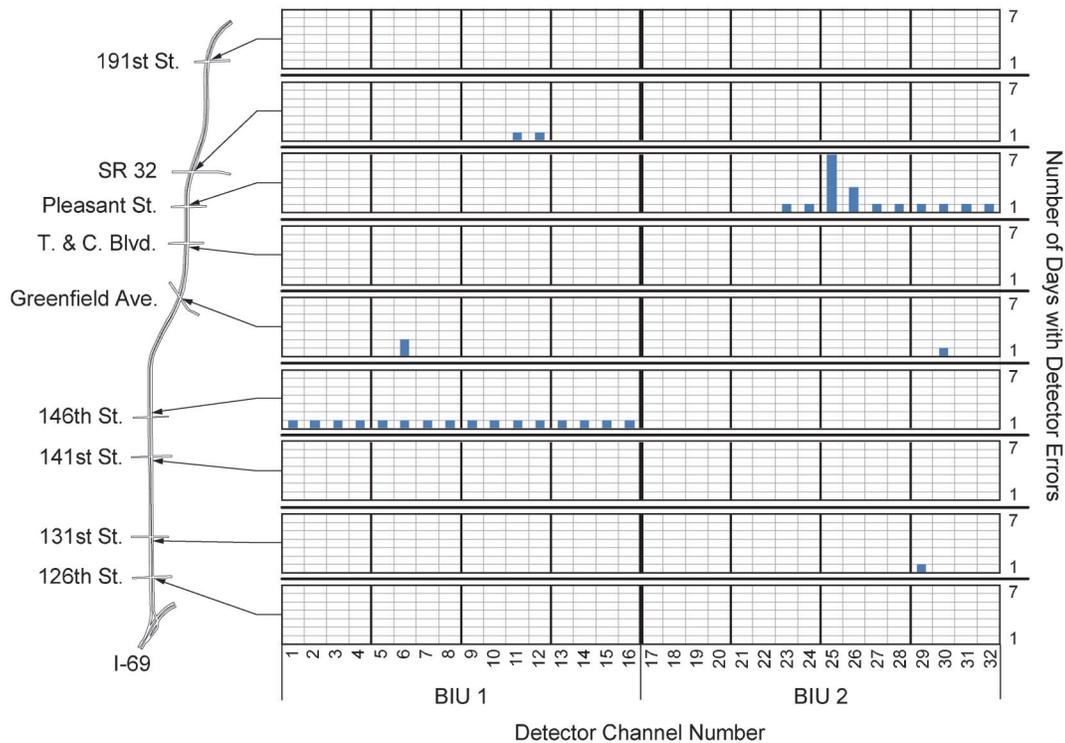


Figure 8.5 Significant detector errors per 24-hour period for an arterial (SR 37 in Noblesville, Indiana) (15).

with a time-stamped record of the beginning and end of an error. The integration of a data logging system into a signal controller can make this possible. A number of detector error events are available in the current version of the data specification (see Table 4.6). From these data, it is possible to aggregate the occurrence of errors across time and location.

Figure 8.5 shows a view of the prevalence of detector errors by intersection for the first two bus interface units (BIUs), which correspond to detector channels 1–32. For each channel number at each intersection, the number of days in one week where an error occurred is plotted. For example, channel 25 at the Pleasant Street intersection had an error every day of the week. There also appears to be a transient rack or BIU failure at 146th Street on one particular day. There are other sporadic errors occurring at various locations, but most of these are not severe. There are also four intersections with no logged errors during this particular week. Most likely, the only site that warrants a field visit is Pleasant Street, because of the consistent error occurring on channel 25. Channel 26 may also warrant inspection during a technician visit to the location.

9. OUTCOME ASSESSMENT

This chapter discusses alternative uses of travel time data for assessing traffic signal operations. This includes a case study in which the outcome of a signal retiming project is assessed using travel time data. This discussion is adapted from recent papers by Remias *et al.* (72,73).

9.1 Travel Time Data Sources

In recent years, a number of technologies have emerged enabling travel time information to be collected on public roads (72–88). The probe data market is not yet mature enough to be able to effectively measure minor movements and side street delay, except where volumes are exceptionally high. However, it is currently possible to assess arterial operations using these data sets. The large numbers of samples obtainable from the new travel time data sets support more sophisticated analysis techniques than those intended for floating-car studies. These data are highly variable because of the complexity of the control system, with traffic patterns along a signalized corridor having many entry and exit points and many locations where motorists make brief detours from their journeys to visit shops and other businesses along the roadway.

There are several existing probe vehicle technologies, each having its own advantages and disadvantages.

- *Agency-driven probe vehicles.* Agency vehicles are driven through a corridor with GPS units to measure travel times. This provides a high-resolution record of one particular vehicle journey, but an extremely small number of actual travel times.
- *Vehicle re-identification using pavement sensors.* Sensors are placed in the pavement to detect a vehicle “signature” (the sensor response to the passing vehicle), which can be matched against other sensors on the roadway, or other roadways (76). This methodology can provide a large number of samples, but generally requires permanent sensor installation.

- *Vehicle re-identification using MAC address matching.* Bluetooth monitoring stations that collect MAC addresses from cellular phones, GPS units, and other electronic devices are located at various locations along a corridor. The MAC addresses can be matched from station to station to calculate travel times (77–81). This provides a lower number of samples than from permanent sensors, but there is greater flexibility in setting the locations of the monitoring stations.
- *Crowdsourced data.* Using mobile phone applications and commercial GPS data, vehicle trajectories along roadway segments are tracked and aggregated per minute to provide vehicle travel times on pre-defined highway map segments (83,84). These data have only recently become available to agencies. This type of data is referred to as “crowdsourced” as it relies on individual road users to “opt in” and contribute to the data set (in exchange for travel information).
- *Virtual probe model.* This technique uses high-resolution event data at each intersection along an arterial to estimate probable vehicle trajectories along the corridor. This yields an estimate of the arterial travel time (85–87).
- The *median* (where the CFD crosses the 0.5 line) is a reasonable measure of central tendency for many distributions.
- The *slope* of the CFD curve relates to the degree of variability in the travel time data, and thus the reliability of travel time. A steeper slope indicates a smaller range, and less variability (higher reliability), while a shallower slope indicates a wider range and more variability (lower reliability). The interquartile range, or the difference between the 75th and 25th percentiles, characterizes the degree of variability, generally without excessive influence from prominent tails in the distribution.
- *Changes in slope* are commonly seen in travel time distributions for signalized facilities. These bends in the CFD indicate modal tendencies in the distribution, which relate to the prominence of vehicle stops along the route. Vehicles that are stopped at an intersection will have their travel times increased by an amount related to the cycle length—generally no less than the minimum amount of time needed to serve crossing movements. Split failures and multiple stops will tend to increase this by integer multiples of the cycle length, causing the stratification as seen in Period 1 in Figure 9.1a.

This chapter focuses on vehicle re-identification (using wireless magnetometers and MAC address matching) and crowdsourced data to assess arterial traffic signal operations. Case studies empirically demonstrate how travel time data can identify different aspects of arterial operations.

9.2 Data Processing and Analysis

On the US 36 test corridor in Indianapolis, Indiana (Figure 4.2), travel times through the corridor were analyzed using the Wireless Magnetometer Vehicle Re-identification System (WMVRS) (76). Nine magnetometer arrays were installed along the corridor, creating six paths for travel time measurement (called “links”).

Figure 9.1a plots individual travel times for a westbound lane between Post Road and Franklin Road on US 36. The westbound direction experiences heavy congestion in the a.m. peak period, as many commuters are en route from the suburbs to downtown Indianapolis. In Figure 9.1a, the a.m. peak travel time data exhibits three modes. Callout “i” in Figure 9.1a points out vehicles traveling at the free flow speed along the segment. Callout “ii” shows vehicles that stopped at one traffic signal. Callout “iii” shows vehicles that stopped at multiple signals or experienced a split failure at one intersection. These modal groupings represent opportunities for improvements to signal timing such as adjustments to splits or offsets.

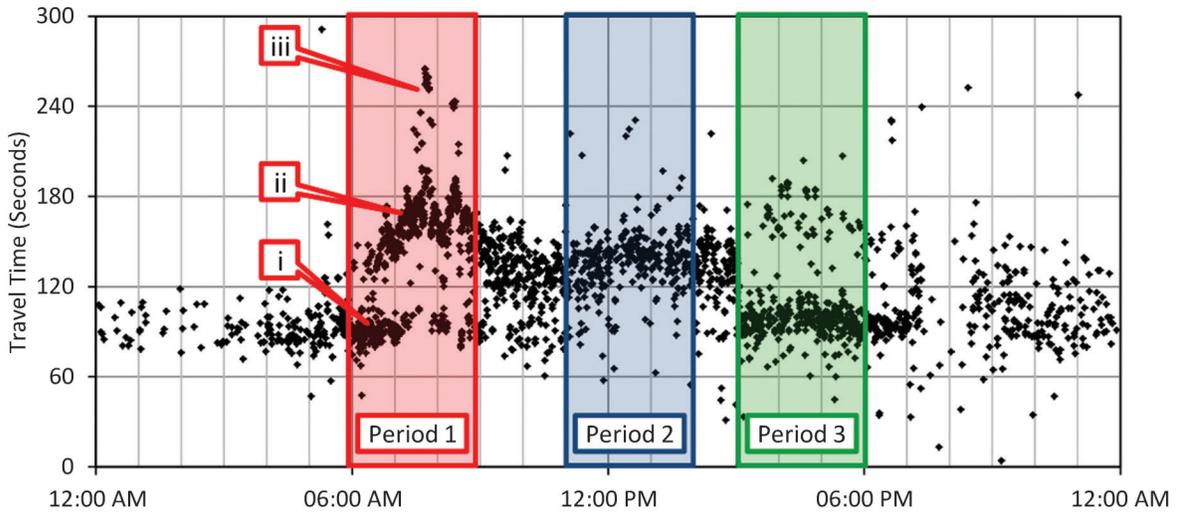
Another interesting feature of Figure 9.1a is the three distinct travel time patterns among the three periods marked periods 1, 2, and 3. These three characteristically shaped point clouds correspond to three time-of-day plans in the signal control. The points in these time periods can be represented graphically as histograms, as in Figure 9.1b–d. These data can also be viewed as cumulative frequency diagrams (CFDs), shown in Figure 9.1e–f. Several properties of the CFDs relate to characteristics of travel time and travel time reliability:

The travel time data enable the analyst to see where opportunities exist to make improvements. Figure 9.2 overlays the three CFDs from Figure 9.1a, Figure 9.1b, and Figure 9.1c (shown respectively using callouts “i,” “ii,” and “iii” in Figure 9.2). It is clear that line “i” has the best median travel time. However, this distribution moves sharply to the right, near the 85th percentile mark. Line “ii” has the best overall reliability, but not the best median travel time. Line “iii” has poor reliability and the worst median travel time. Visualizing travel times with CFDs is valuable for comparing both travel time and travel time reliability.

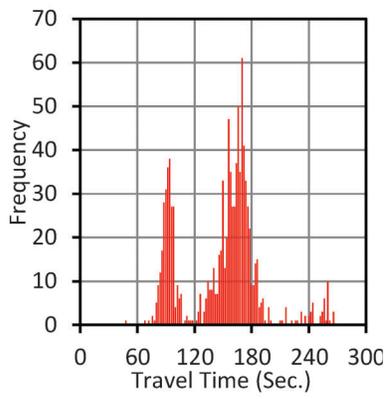
9.3 Comparison with Crowdsourced Data

The deployment of sensor equipment for vehicle re-identification requires agency investments that might not always be feasible. Newer data sources are currently emerging for analyzing corridor operations. Of particular interest is the development of crowdsourced data that rely on mobile electronic devices, especially mobile phones, but also GPS devices and other vehicle navigation systems. INRIX currently offers 1-minute average travel times for predefined highway segments (called Traffic Message Channels, or “TMCs”). Unfortunately, the base maps for these data are sometimes incongruous with the boundaries of traffic signal systems.

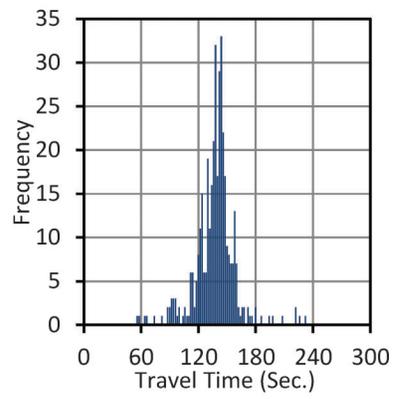
Crowdsourced data were examined for the US 36 corridor in Indianapolis, Indiana (Figure 9.3a). Travel times along 3.8 miles of the corridor (Figure 9.3b) were evaluated to determine potential timing plan improvements or opportunities to deploy advanced technology such as adaptive control. Figure 9.3c and Figure 9.3d show CFDs of half-hour distributions of travel time for each weekday in a month. Figure 9.3c shows the 11 a.m. hour for June 2012 and Figure 9.3d shows the 5 p.m. hour. Because of the low density of



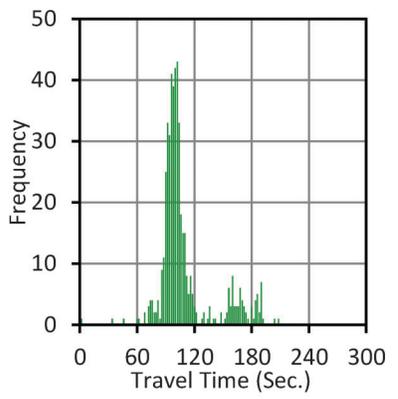
(a) Scatter plot of travel times over a day



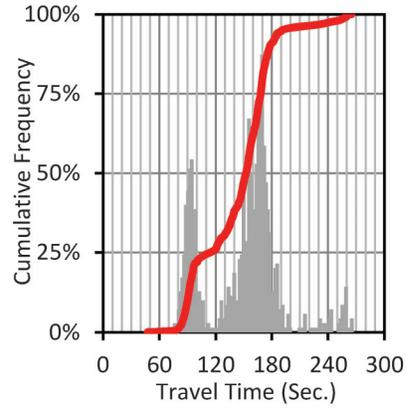
(b) Period 1 histogram



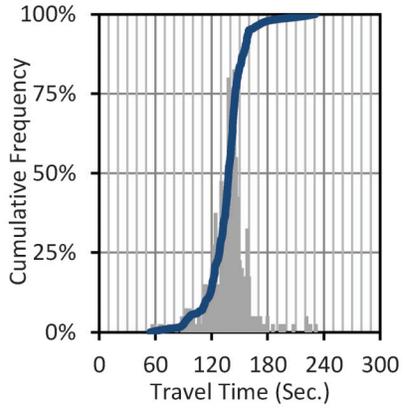
(c) Period 2 histogram



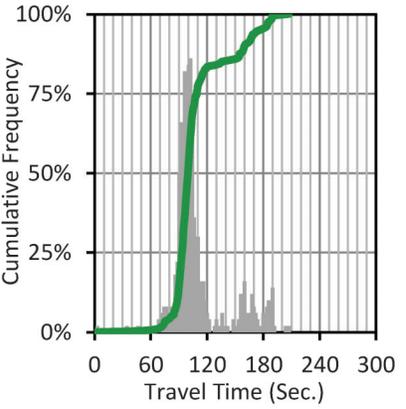
(d) Period 3 histogram



(e) Period 1 CFD



(f) Period 2 CFD



(g) Period 3 CFD

Figure 9.1 Travel times: Westbound, US 36.

the crowdsourced data on arterial streets, only 38 of 40 possible half-hour increments were obtainable from the source data. As the rate of vehicles participating in crowdsourcing increases, the fidelity of these plots should substantially increase.

Figure 9.3c shows a relatively consistent travel time distribution during the 11 a.m. hour over the month of

June, with the exceptions of callout “i” and callout “ii.” These two thicker lines represent half hours where there was a short-term disruption, and there may exist opportunities to improve the system. In comparison, Figure 9.3d represents the half hour CFDs for the 5 p.m. hour over the month of June. As expected, the opportunities where improvements could be made are

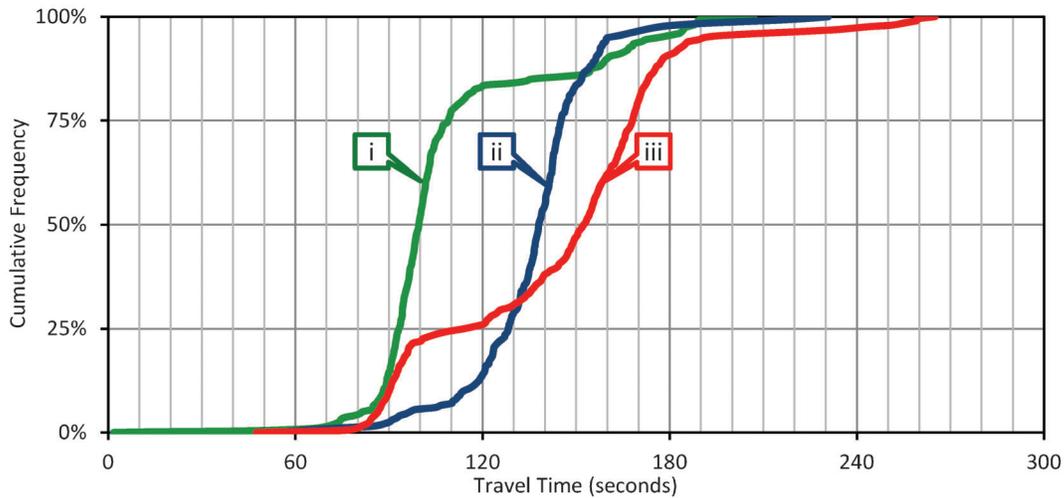


Figure 9.2 US 36 Travel time CFDs for three periods during a weekday.

far greater during the 5 p.m. hour, and are shown as the thick red lines. Figure 9.3c and Figure 9.3d provide an example for identifying when signal timing changes could have a positive impact.

9.4 Sensor Placement and Data Reduction Techniques

While corridor travel time is an appealing metric, especially for coordinated signal systems, identifying the relative delay at each signal is also important. Figure 9.4 shows three different ways to characterize travel time segments along an arterial. The example roadway is a commuter corridor connecting Indianapolis and Bloomington, Indiana. The same data from each Bluetooth monitoring station (BMS) is used for all three segmentation strategies, and three different stories emerge from the data. These are all useful, depending on the user objective. The reduction techniques are origin-based travel time, destination-based travel time, and subsegment travel time.

9.4.1 Origin-Based Travel Time

The first data reduction technique is called “origin-based travel time” (Figure 9.4a and Figure 9.5a). During the p.m. peak period from 16:00 to 19:00, the workforce from Indianapolis leaves the city and the southbound traffic is quite heavy (AADT exceeds 30,000 vehicles per day). The northernmost BMS (labeled “BMS-1” in Figure 9.5a) is where each of the origin-based travel time segments originates. Thus, a family of sequentially larger travel time segments is formed with BMS-1 to BMS-2, BMS-1 to BMS-3, BMS-1 to BMS-4, and so on. In current practice, often only the average travel time is examined along the entire corridor between two prominent endpoints. The use of averages and standard deviations to describe large sample sets is a legacy of times when modern computers and databases capable of handling large data sets were not readily available.

Instead of assuming a normal distribution and calculating a standard deviation, the distribution for

each travel time segment is plotted in a cumulative distribution of the field-collected sample, as shown in Figure 9.5. If we consider only the mean corridor travel time (as in current practice), the corresponding data would be the average travel time between BMS-1 and BMS-8. However, a visual examination of the CFDs in Figure 9.5 reveals that the slope generally decreases as we scan from left to right across the figure. This corresponds to an increase in the interquartile range, and shows that the reliability of the system degrades as we increasingly widen the distance between the endpoints for the travel time measurement. This is as expected, since each additional traffic signal represents an additional opportunity for a stop.

9.4.2 Destination-Based Travel Time

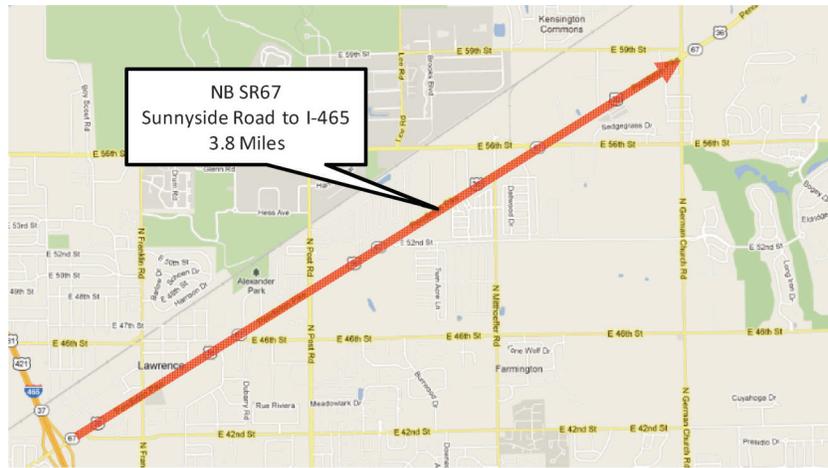
The next data reduction technique, shown in Figure 9.4b is “destination-based travel time.” Since the movement of concern is the southbound through traffic, most of the vehicles proceed through the system and pass the southernmost BMS (BMS-8). Although some cars exit the arterial north of BMS-8 and others enter the system south of BMS-1, it is still important to examine how the southern portion of the corridor operates. The curves of the destination-based cumulative frequencies shown in Figure 9.5b look quite similar to the origin-based travel time cumulative frequencies curves shown in Figure 9.5a. However, as discussed in the following section, a slight anomaly is discernible between BMS-5 to BMS-8 and BMS-4 to BMS-8.

9.4.3 Subsegment Travel Time, or Control Delay

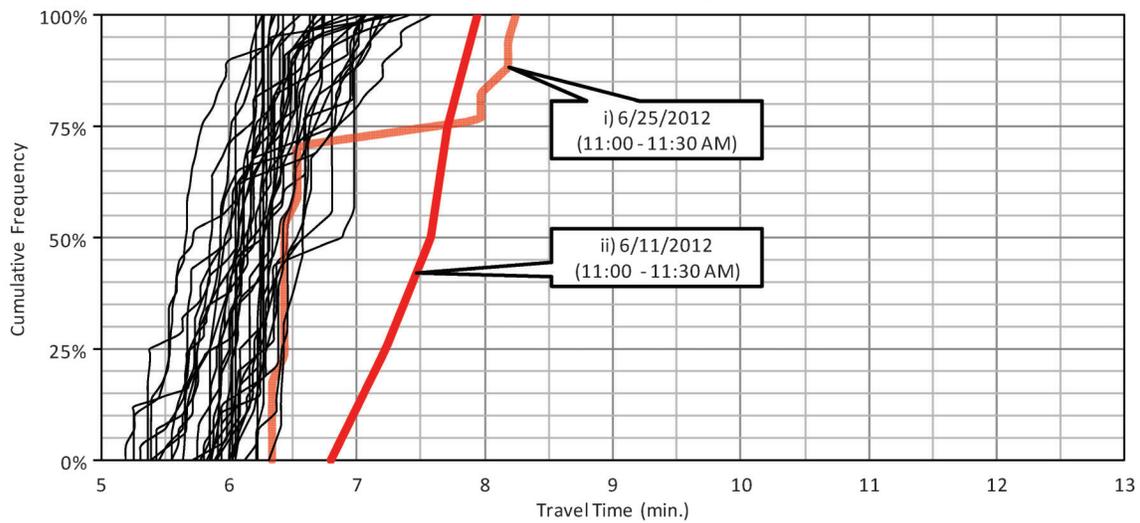
The third data reduction technique, shown in Figure 9.4c, is called “individual intersection approach delay segmentation.” This method examines the southbound traffic on individual links within the corridor. This creates a series of links including BMS-1 to BMS-2, BMS-2 to BMS-3, BMS-3 to BMS-4, and so forth.



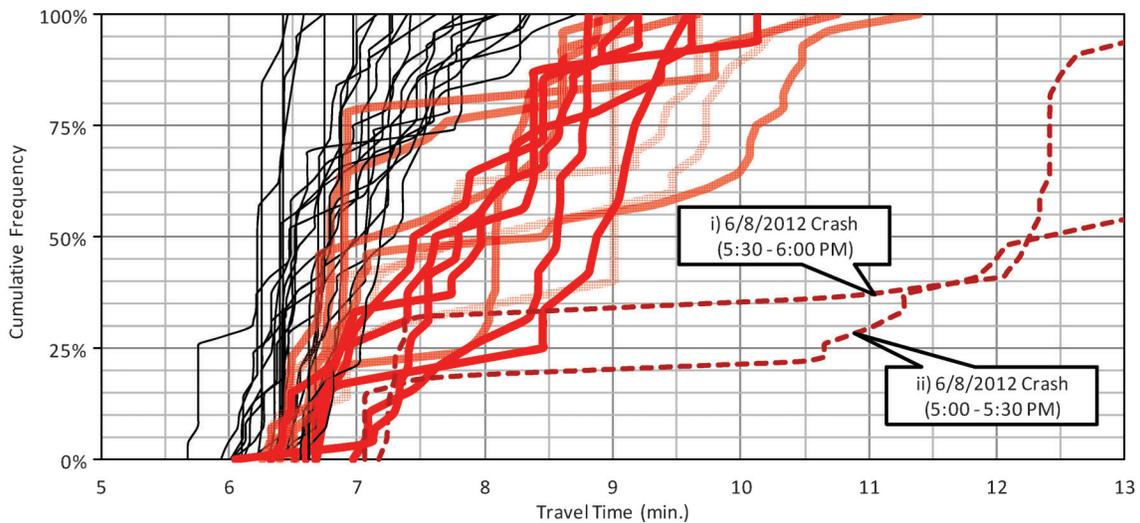
(a) US 36 Location.



(b) US 36 Corridor.



(c) Half-hour travel distributions using crowdsourced data for the 11 a.m. hour



(d) Half-hour travel distributions using crowdsourced data for the 5 p.m. hour

Figure 9.3 June 2012 weekday travel time distributions for 11 a.m. and 5 p.m. hours.

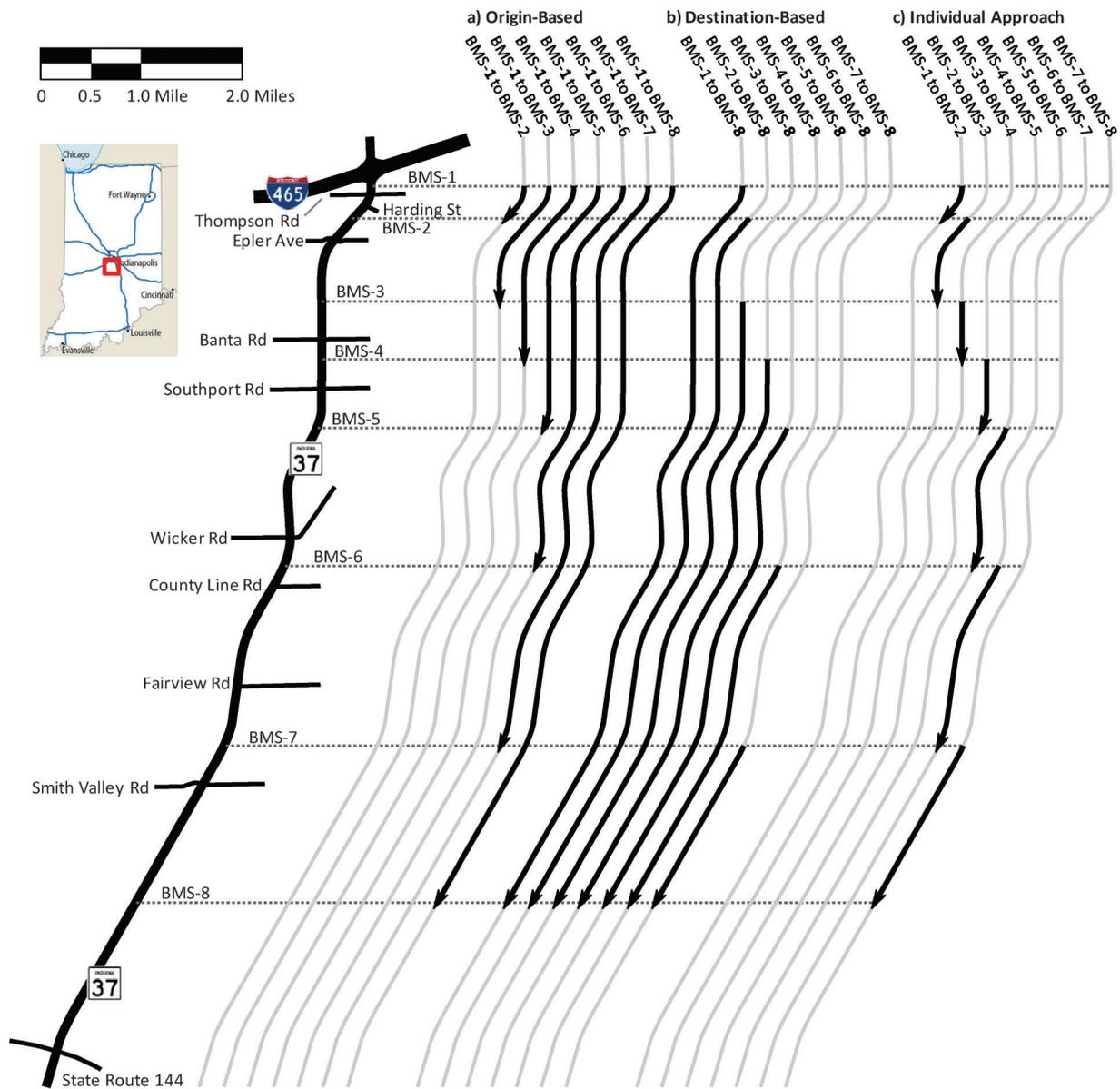
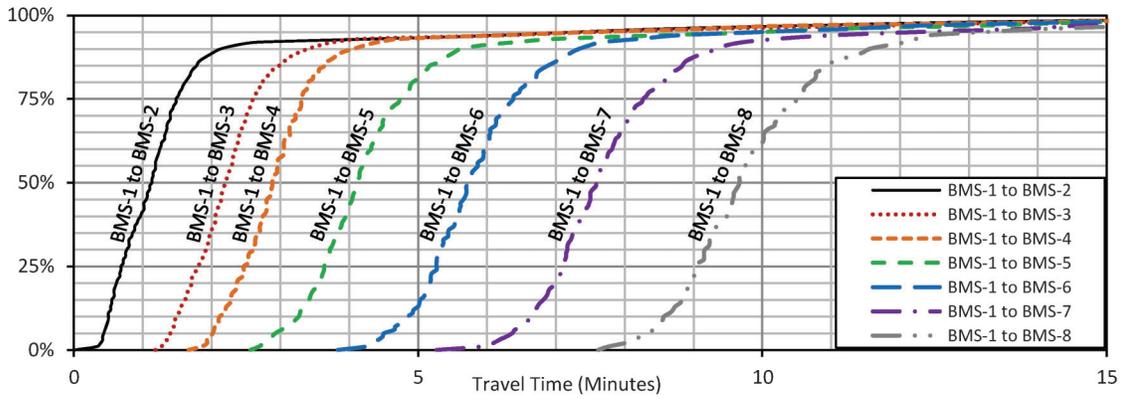


Figure 9.4 Bluetooth monitoring stations and subsection segment regimes on SR 37 south.

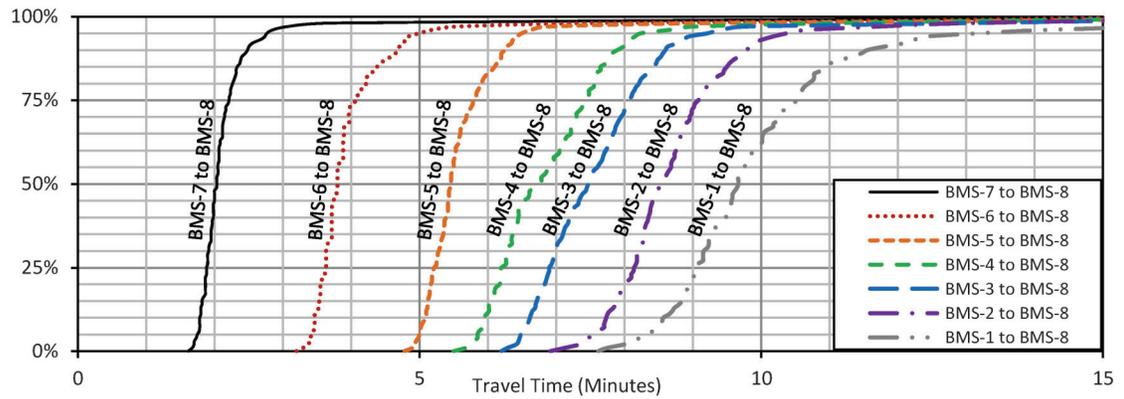
Because the length of each link varies, control delay rather than travel time will be estimated, to better compare the results from each link. Control delay is the travel time minus the time required to proceed through the segment at the free flow speed. The 5th percentile travel time was selected to define the free flow speed. Figure 9.4c shows the cumulative frequencies for each sample after the travel times are converted to control delays for each subsegment.

The curve labeled “i” for BMS-1 to BMS-2 (corresponding to the same curve in Figure 9.5a, but shifted left by the 5th percentile travel time to obtain control delay) and the curve labeled “ii” for BMS-4 to BMS-5 both clearly appear as curves that should be further examined.

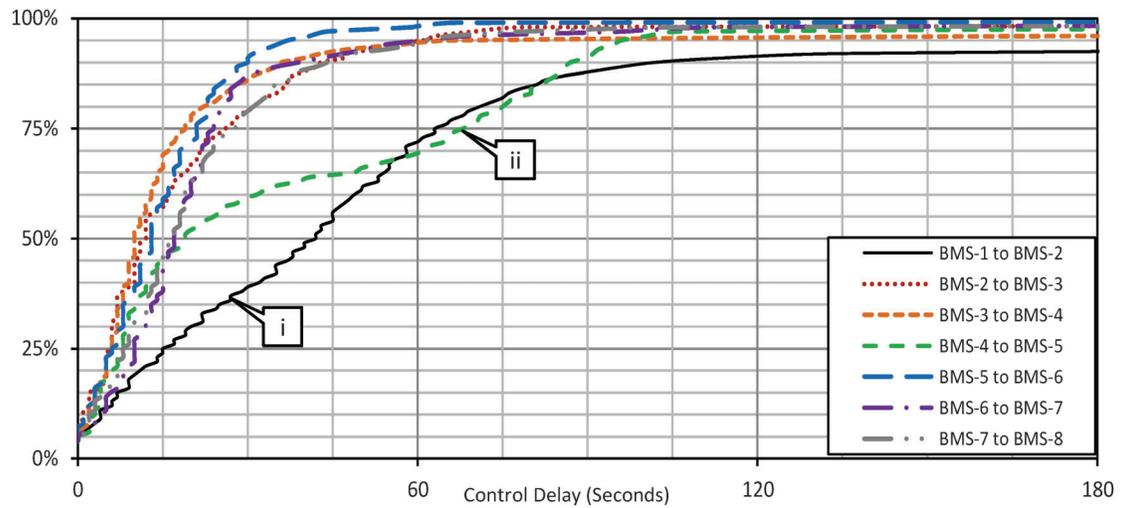
Curve “i” shows the same shape as in Figure 9.5a. This subsegment is near the exit ramp of I-465 and experiences a high number of randomly entering vehicles originating from the freeway. Curve “ii” is more interesting; this curve represents a subsegment in the middle of the corridor. Recall that in the destination-based travel time, the discontinuity in the slope between the 25th and 75th percentiles was an anomaly, and was difficult to distinguish. It is much more apparent in the control delay analysis (Figure 9.5c). This shows a bimodal distribution of travel time, which is experiencing an issue with the only signal in that subsegment (the intersection of SR-37 and Southport Road, Figure 9.4). Although travel time may not be able to explicitly identify the cause of the anomaly,



(a) Origin-based travel time (minutes)



(b) Destination-based travel time (minutes)



(c) Segment control delay (seconds) based on 5th percentile free flow travel time

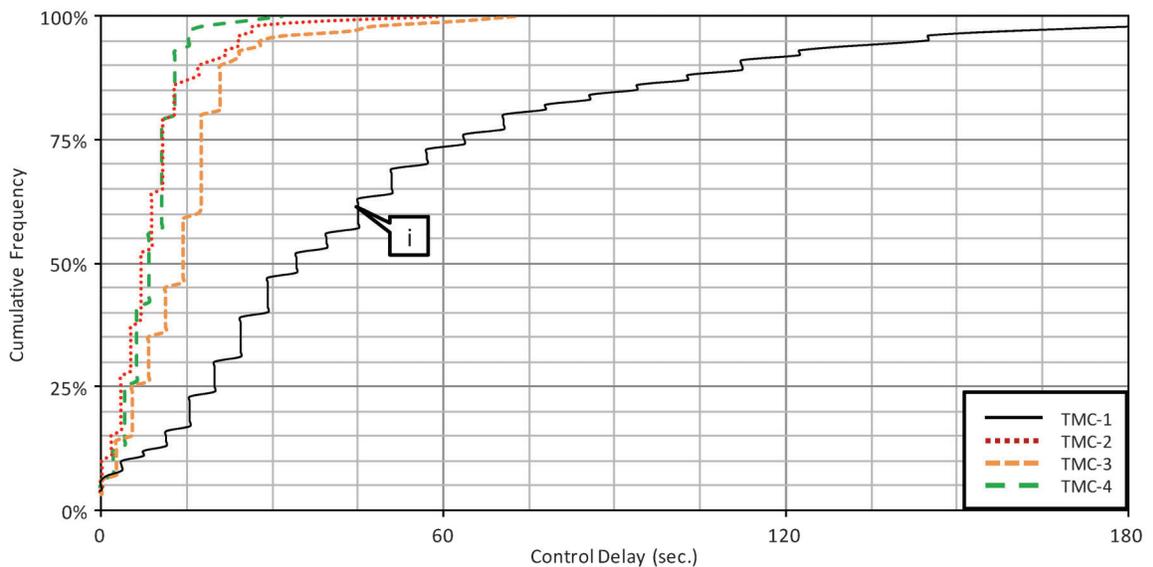
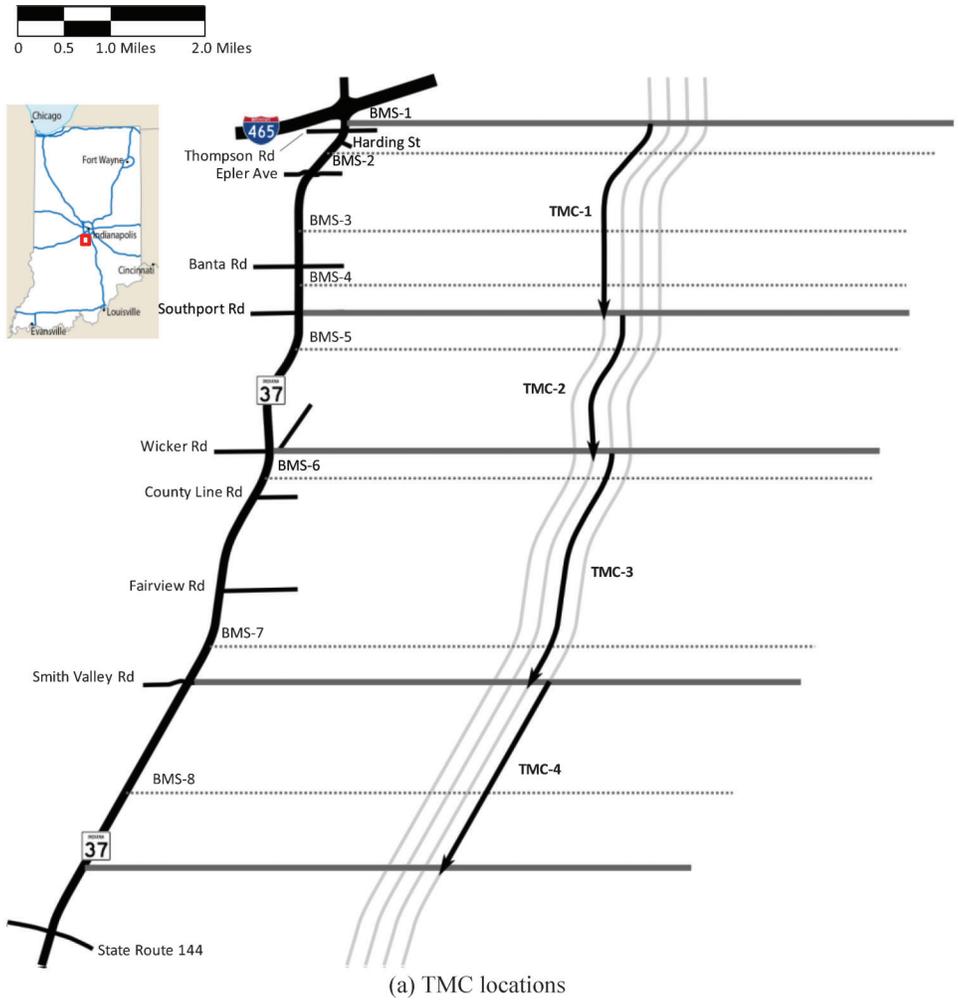
Figure 9.5 Southbound SR 37 travel time analysis during the p.m. peak period.

the characteristics of this curve suggest there is a potential for improvement, such as by changing a signal offset (88).

9.5 Commercial Probe Data

As mentioned earlier, crowdsourced probe vehicle data (in this case from INRIX) consist of aggregated data for predefined roadway sections called TMCs.

The TMCs for SR 37 are illustrated in Figure 9.6. There are four TMCs, which is a coarser segmentation than obtained from the seven BMSs as shown in Figure 9.4. However, similar data reduction techniques can be employed to evaluate the system with the crowdsourced data. Control delay was plotted using the crowdsourced data (Figure 9.6b), illustrating similar trends as seen in the BMS travel times.



(b) SR 37 p.m. Peak hour control delay (seconds) by TMC segment

Figure 9.6 SR 37 corridor control delays using crowdsourced data.

Callout “i” in Figure 9.6b corresponds to callouts “i” and “ii” in Figure 9.5c. The smaller segments of the BMS data allow for better fidelity in locating operational problems. However, as the quality of crowdsourced data continues to increase because of advances in technology and an increasing number of users, these data will serve as a substantial tool for the evaluation of signal system performance.

9.6 Cost Assessment of a Corridor Improvement

In addition to providing an assessment of corridor signal operations, multiple measurements from different time periods support a quantitative assessment of operational changes, such as a signal retiming project. This section provides a case study from an 8.7-mile corridor in Kokomo, Indiana (Figure 9.7b and Figure 9.7c).

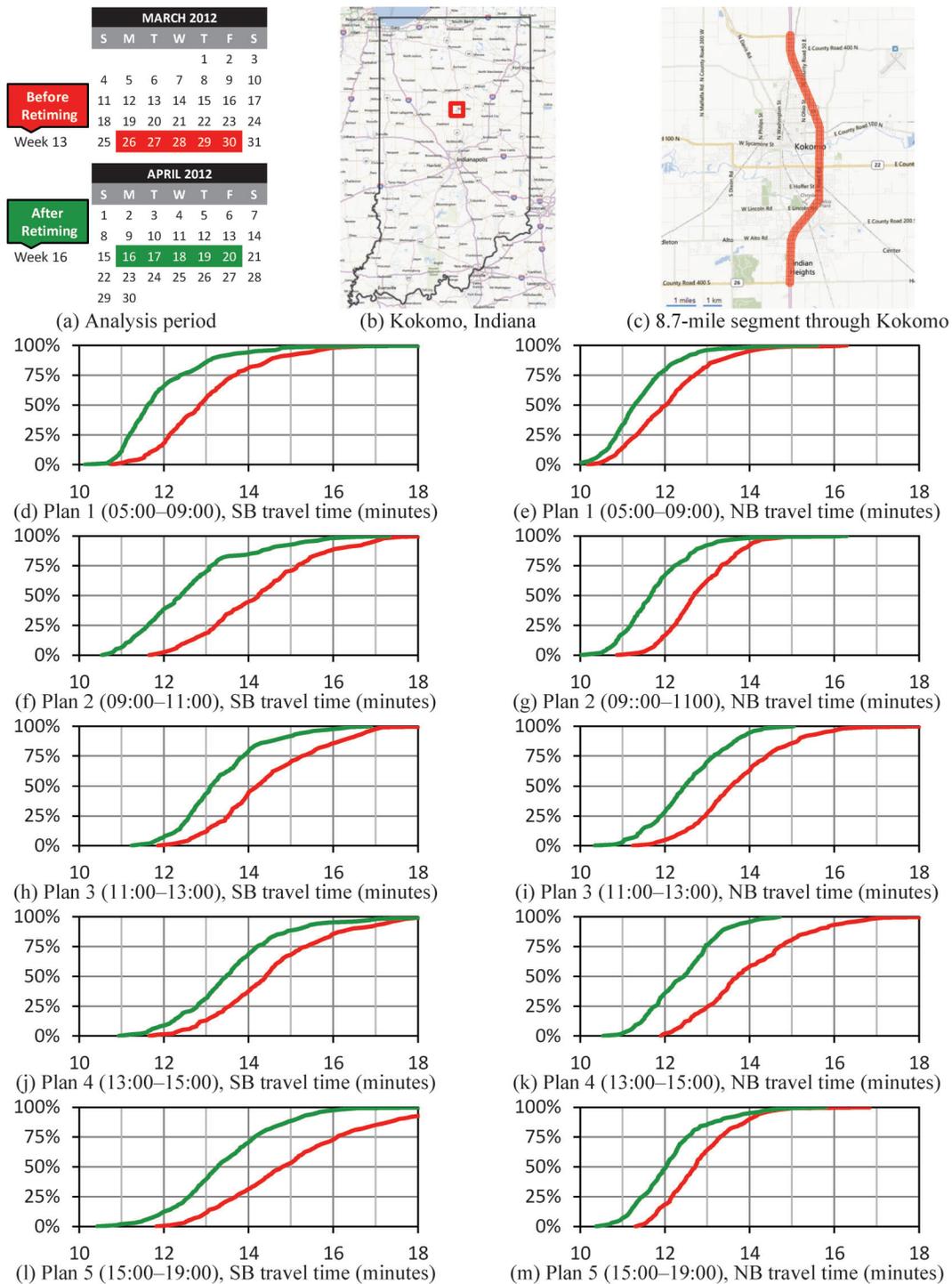


Figure 9.7 Travel time data from US 31 retiming in Kokomo, Indiana.

A signal retiming project was executed on this corridor during April 2012. For the purpose of data selection, the week before the retiming (week 13 of 2012) was defined as the “before” period and the week after the retiming (week 16 of 2012) as the “after” period (Figure 9.7a). Northbound and southbound travel times were obtained for each of the five time-of-day plans (Figure 9.7d–m). In every case, travel times were reduced. Because US 31 primarily serves through traffic, the goal of reducing the through-movement travel time was achieved. On other arterials, a different objective (such as reducing overall intersection delay) might be more appropriate—especially if there is substantially more local traffic than through traffic.

A calculation can be performed to quantify the benefits of traffic signal retiming. Using the median travel times throughout the different weekday timing plans, Table 9.1, and the average annual daily traffic (AADT) of 26,000 vehicles on the corridor, a monetary value can be associated with the benefit. Currently there are numerous strategies for quantifying travel time savings. Here, the Texas Transportation Institute (TTI) methodology is used (82).

First, the change in travel time (ΔTT) is calculated:

$$\Delta TT = TT_{\text{before}} - TT_{\text{after}}, \quad (9.1)$$

where TT_{before} is the travel time before retiming and TT_{after} is the travel time after retiming. The travel time is measured in minutes, and a different value is obtained for each time-of-day plan and for each individual direction (northbound and southbound). Costs for trucks are given by

$$COST_{\text{TRUCK}} = (\Delta TT)(VOL)(P_{\text{TRUCK}}) (PPV_{\text{TRUCK}})(HC_{\text{TRUCK}}) \left(\frac{1}{60}\right), \quad (9.2)$$

where $COST_{\text{TRUCK}}$ is the user cost for trucks, VOL is the number of vehicles measured for the study period, P_{TRUCK} is the percentage of commercial trucks (4% for weekdays), PPV_{TRUCK} is the number of passengers per truck (equal to 1), HC_{TRUCK} is the hourly cost for trucks (equal to \$86.81), and the 1/60 factor converts the travel time change from minutes to hours. The \$86.81 amount represents the commercial vehicle operating costs in 2011 dollars and is taken from the 2011 Transportation Urban Mobility Report (82). This value does not represent excess fuel consumption. When ΔTT is positive, the outcomes of the equation reflect a user savings. Costs for passenger cars are given by

$$COST_{\text{PC}} = (\Delta TT)(VOL)(P_{\text{PC}}) (PPV_{\text{PC}})(HC_{\text{PC}}) \left(\frac{1}{60}\right), \quad (9.3)$$

TABLE 9.1
Median travel time savings on US-31 in Kokomo, Indiana.

	Plan	Median TT Savings (min)	% of Daily Traffic	TT Savings (h)	TTI Travel Time Savings (\$)	CO ₂ Reduction (tons)	CO ₂ Emission Savings (\$)
Southbound US 31	Plan 0 (:00:00–05:00)	0.79	2.2%	1987.34	\$ 46,941.69	16.77	\$ 368.96
	Plan 1 (05:00–09:00)	1.22	7.2%	9925.88	\$ 234,453.24	83.76	\$ 1,842.82
	Plan 2 (09:00–11:00)	1.83	5.3%	10877.93	\$ 256,941.12	91.80	\$ 2,019.58
	Plan 3 (11:00–13:00)	1.1	6.7%	8246.25	\$ 194,779.77	69.59	\$ 1,530.98
	Plan 4 (13:00–15:00)	0.93	6.6%	6886.14	\$ 162,653.47	58.11	\$ 1,278.47
	Plan 5 (15:00–19:00)	1.53	13.5%	23311.22	\$ 550,620.34	196.72	\$ 4,327.91
	Plan 6 (19:00–24:00)	0.91	7.1%	7319.89	\$ 172,898.62	61.77	\$ 1,559.00
Northbound US 31	Plan 0 (:00:00–05:00)	0.58	2.2%	1462.30	\$ 34,540.02	12.34	\$ 271.49
	Plan 1 (05:00–09:00)	0.75	7.6%	6420.27	\$ 151,649.25	54.18	\$ 1,191.97
	Plan 2 (09:00–11:00)	1.02	5.5%	6316.57	\$ 149,199.92	53.31	\$ 1,172.72
	Plan 3 (11:00–13:00)	1.1	7.0%	8627.08	\$ 203,775.18	72.80	\$ 1,601.69
	Plan 4 (13:00–15:00)	1.26	7.0%	9881.93	\$ 233,415.21	83.39	\$ 1,834.66
	Plan 5 (15:00–19:00)	0.69	14.2%	11040.76	\$ 260,787.26	93.17	\$ 2,049.81
	Plan 6 (19:00–24:00)	0.45	7.9%	4018.01	\$ 94,906.91	33.91	\$ 745.97
Total			116321.6	\$ 2,747,562	981.64	\$ 21,596.03	

where $COST_{PC}$ is the user cost for passenger cars, VOL is the same as in the previous equation, P_{PC} is the percentage of passenger cars (96% for weekdays), PPV_{PC} is the number of passengers per vehicle (estimated at 1.25), HC_{PC} is the hourly cost for passenger cars (\$16.79), and the 1/60 factor again converts travel time from minutes to hours.

In addition to user costs, potential savings in fuel consumption and associated changes in carbon dioxide (CO_2) emissions can be derived as follows:

$$FUEL = (\Delta TT)(VOL)(MPG) \left(\frac{1}{60} \right). \quad (9.4)$$

In this equation, $FUEL$ is change in amount of fuel consumed (gallons), which is a saving when ΔTT is positive. With the use of conversion factors from Argonne National Laboratory, a passenger car that idles at 1,000 rpm with air conditioning on 50% of the time can be expected to consume 0.87 gal of gasoline per hour, or 0.0145 gal/min. This number was used to conservatively estimate the change in fuel consumption for all vehicle types associated with changes in travel time (positive ΔTT), the amount of carbon dioxide (CO_2) emissions that is prevented is calculated from the following two equations:

$$CO_2 = (FUEL)(19.4 \text{ lb/gal}) \\ (0.0005 \text{ ton/lb}), \quad (9.5)$$

$$COST_{CO_2} = (CO_2)(\$22). \quad (9.6)$$

Here, $COST_{CO_2}$ represents the CO_2 cost. According to the U.S. Environmental Protection Agency, the amount of CO_2 emitted when a gallon of gasoline burns is approximately 19.4 lb/gal. The monetary equivalent of the CO_2 produced is assumed to be \$22/ton of CO_2 produced.

The results shown in Table 9.1 illustrate the benefit for system users on the basis of the previous analysis. The savings are calculated from changes in arterial travel time measured with crowdsourced probe vehicle data, and volumes are assumed using AADT and an INDOT counting station. With the use of retiming, the US 31 corridor user cost reductions over the entire year exceed \$2.7 million and there is a nearly 1000-ton reduction in CO_2 emissions.

9.7 Signal Retiming Aging

The above calculations were performed using one week of crowdsourced data just before retiming and one week of crowdsourced data immediately after the retiming project was completed. However, the literature is silent on how traffic signal timing plans age, and many agencies often use a two- to three-year cycle for retiming corridors. Crowdsourced data provides a cost effective mechanism for evaluating how a traffic signal timing plan ages. Understanding how a timing plan ages, can be beneficial when adapting agency

standards for how often signal systems should be retimed (73).

9.7.1 Weekly Analysis of Signal Retiming Aging

Figure 9.8 shows 63 weeks of travel times along the US-31 corridor as cumulative frequency distributions for both directions and each of the timing plans. Each line corresponds to approximately 350 and 750 samples. The two dark lines are the original before and after CFDs that correspond to the calculated benefits summarized in Figure 9.7, and the thinner lines correspond to the other 61 weeks. The red lines are weekly travel time summaries between January 1 and March 30, 2012, before the corridor was re-timed, and the green lines are weekly travel time summaries between April 16, 2012, and March 31, 2013, after the corridor was retimed. The pattern of improvement from the 13 weeks before the retiming and the 50 weeks after the retiming remains relatively steady, with few outlier weeks. Those outliers are not clearly understood, but are most likely due to construction, crashes, severe weather, or special events.

9.7.2 Monthly Analysis of Signal Retiming Aging

Figure 9.8 suggests that weekly travel time distributions do have some outlier periods that may need to be filtered. Monthly cumulative frequency distributions may be used to provide a clearer picture of long-term aging and seasonal variation trends. In Figure 9.9, each of the lines in the monthly CFDs accounts for between 1,500 and 3,000 travel times, depending on the length of the timing plan period. This is a significant amount of data, especially considering that a majority of before/after signal retiming studies are based on just a few probe vehicle data points from driving through the corridor. Figure 9.9 shows the monthly cumulative frequency distributions for the three months before the retiming and the 11 months following the retiming (April was excluded because it was the month of the retiming). The thick lines are again the original before-and-after week summary from Figure 9.7 for context. The improvement is clear for the 11 months following the retiming of the corridor.

The median monthly travel times for each of the signal timing plans are another way to visualize the improvement of the corridor and to see if there is any rebound in the plan. Figure 9.10 shows 15 months of median travel times for timing plan 4 (13:00–15:00) for the northbound and southbound directions of the corridor. The southbound direction in Figure 9.10a has a median travel time reduction between 0.7 and 1.48 minutes. Comparing the months of March from 2012 (before the retiming) and 2013 (after the retiming), there was a 1.16-minute reduction in travel time. This shows that over the course of the year the retiming did not show any significant aging. The US-31 Northbound median travel times showed a similar pattern of

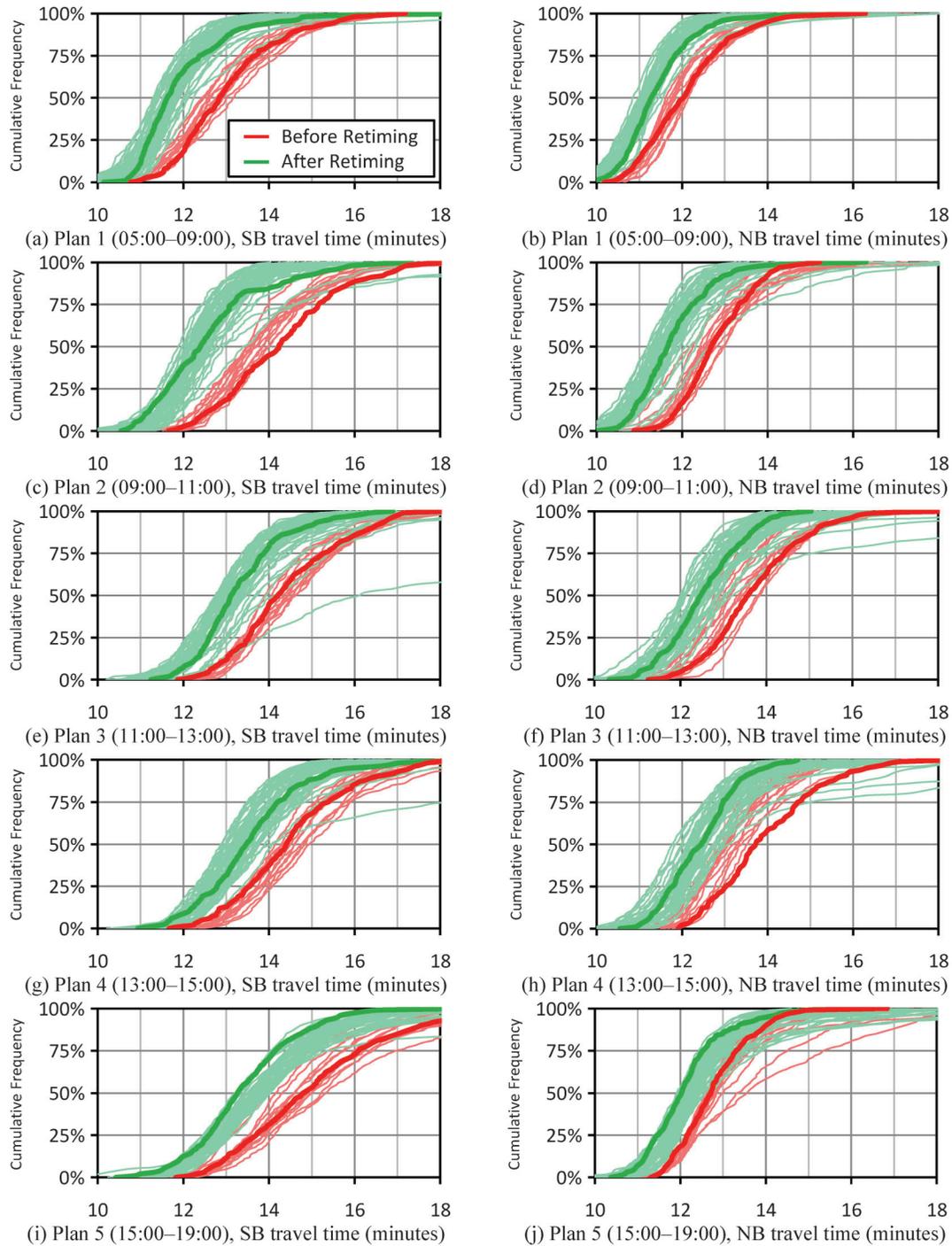


Figure 9.8 Aggregated weekly travel times over 16 months.

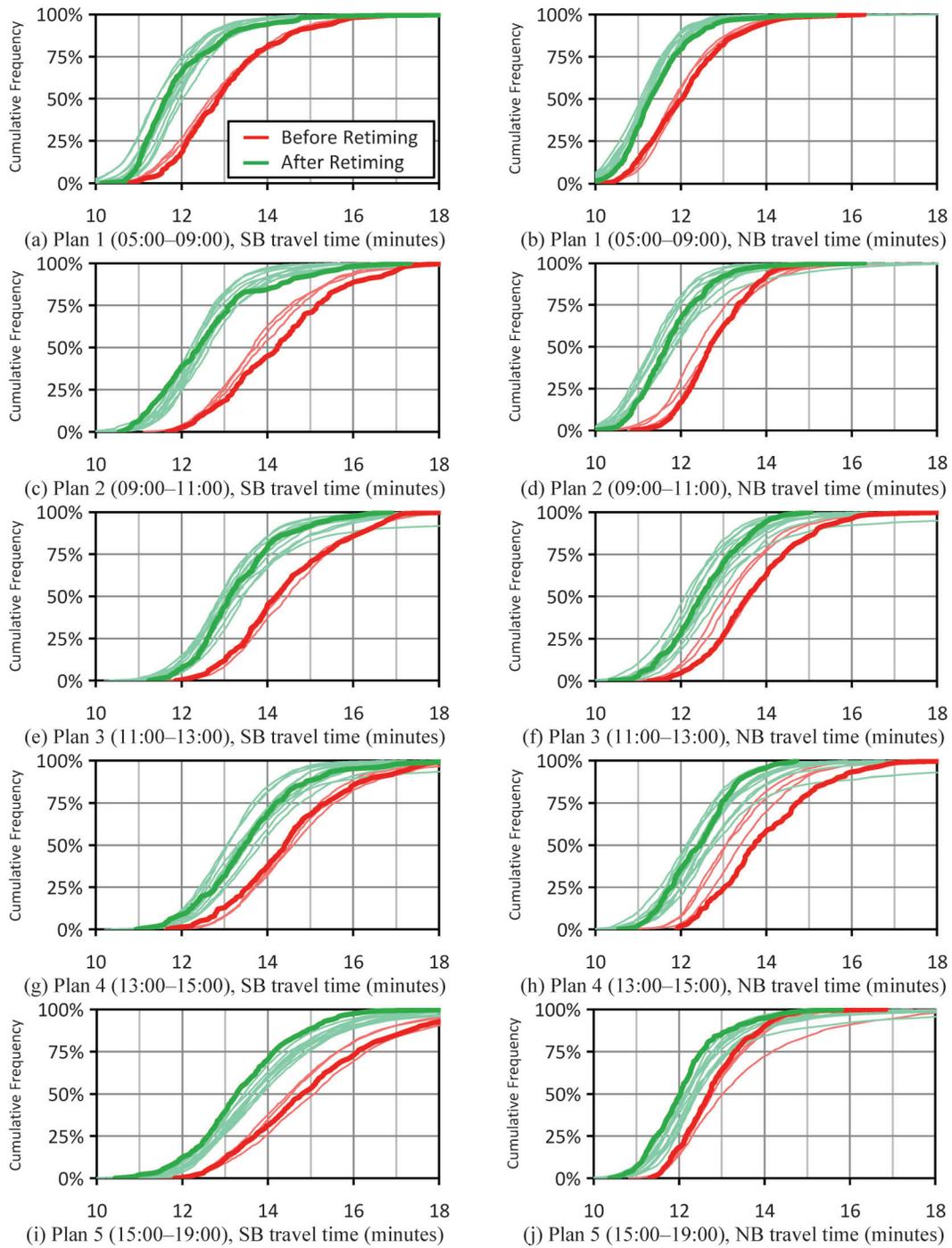
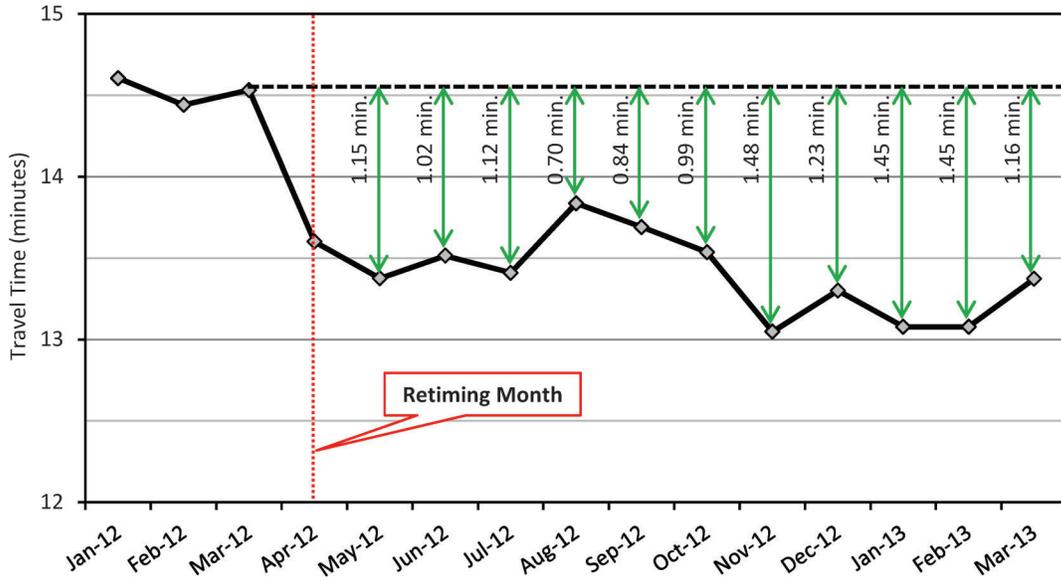
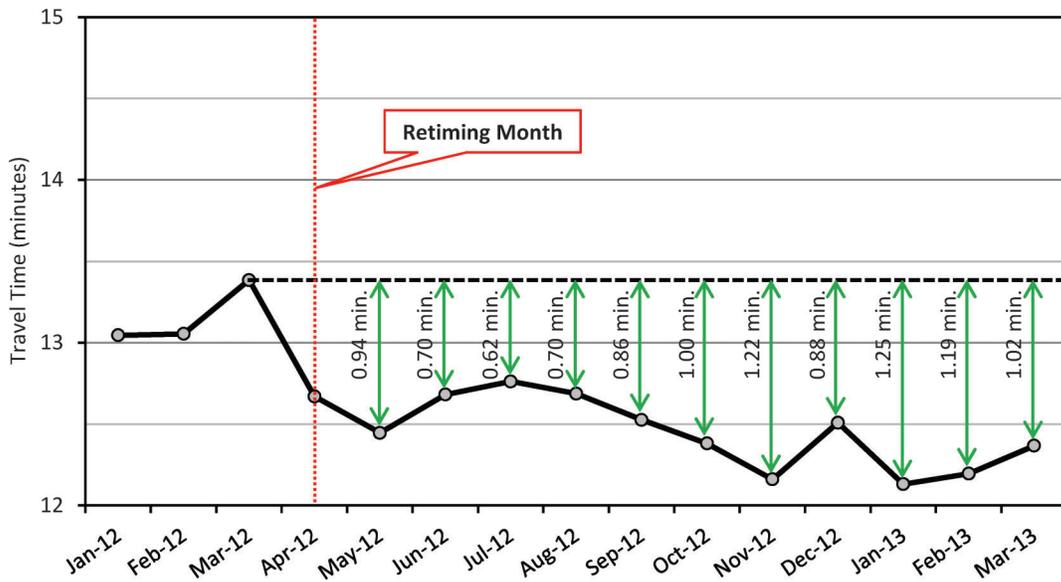


Figure 9.9 Aggregated monthly travel times over 16 months.



(a) Southbound



(b) Northbound

Figure 9.10 Timing plan 4 (13:00–15:00) median monthly travel times.

reduction in travel time in Figure 9.10b. The reduction in median travel time ranged from 0.62 to 1.25 minutes in the northbound direction. There was a 1.02-minute reduction in median travel time from the month of March before the retiming to the month of March after the retiming. The pattern of reduction of median travel time remains constant for both directions and each of the timing plans.

10. SUMMARY

10.1 Overview of Signal Operations and High-Resolution Data

This monograph has provided an overview of signal operation activities and the necessary methodology for collecting data to analyze the system. Chapter 2 provided a discussion of the necessary infrastructure requirements for making such an analysis possible. Chapter 3 discussed various aspects of signal timing as a basis for understanding how to analyze system performance. Chapter 4 presented the methodology for recording and collecting data from the system to be able to subsequently perform the analysis.

10.2 Performance Measures

Chapters 5 through 9 discussed various types of performance measures for traffic signal systems. Chapters 5 and 6 focused on the vehicle mode, with Chapter 5 presenting a set of performance measures focusing on capacity allocation (local control) and Chapter 6 focusing on vehicle progression (system control). Chapter 7 presented multimodal performance measures, including metrics relevant to pedestrian operations, and the handling of preemption and priority requests by the signal controller. Chapter 8 presented performance measures relevant to signal maintenance, including the quality of communications and data collection as well as detector failures. Chapter 9 investigated travel time data on signalized arterials, and presented a case study where the travel time savings from a retiming activity characterized the user benefit.

10.3 Future Opportunities

The methodology presented here is relatively simple to deploy, especially at intersections where the existing detection is already established for actuated-coordinated operations. At the time of writing, the performance measure methodology was deployed at approximately 100 intersections in the state of Indiana. Deployment required an investment in wireless data plans for remote data collection. At some locations, new signal controllers were purchased. However, most of the existing controllers were reaching the end of their service lives, and would have needed replacing regardless of the performance measure initiative. In Indiana, the existing detection could be used at most locations. Estimated deployment costs of

high-resolution data enabled controller and a wireless router are outlined in Table 10.1.

The cost of deploying an adaptive control system is approximately \$65,000 per intersection (14). Such systems are usually procured for one corridor at a time, so procurement for an arterial can be in the hundreds of thousands, or millions of dollars. Such costs are on the scale of capital projects. For a relatively small marginal cost, high-resolution data collection could rather easily be deployed within most projects of this type, and they would provide an independent means of validating, monitoring, and evaluating the control system by establishing a record of the activity at each intersection.

If large procurements are not feasible for an agency, high-resolution data collection can still be incrementally implemented at a cost that may be more compatible with routine maintenance and equipment change-out. Incremental deployment of signal performance measures could enable more rapid improvement of signal operations in many existing situations, in comparison with the slower pace by which larger systems can be procured and installed. The data obtained by these deployments could substantially accelerate retiming activities or enable many of these to be done in a more automated process. The savings in engineering resources would enable agencies to focus on segments of the system requiring the most attention. Furthermore, high-resolution data would enable those priority segments to be identified based on actual system performance in addition to the presently existing tools (engineering judgment, complaint calls, and time elapsed since previous retiming).

As more engineers and researchers obtain access to high-resolution data and begin to exploit the information it contains, it is likely that new applications for the data will emerge. Future research will continue to refine the performance measure methodology and develop case studies to demonstrate its use for improving system operations.

TABLE 10.1
Estimated cost of deploying controller-based high resolution data collection.

Item	Cost
One-Time Costs	
•New Traffic Signal Controller	\$2000
•Wireless Modem	\$800
•Hardened Switch	\$120
•Installation Labor (2 hours at \$100 per hour)	\$200
Total of One-Time Costs	\$3120
Annual Costs	
•12 months of data-only wireless service	\$420/year
Total of Annual Costs	\$420/year
Ten Year Lifetime Cost	
•One-Time Cost	\$3120
•Annual Costs × 10 Years	\$4200
Total Ten-Year Lifetime Cost	\$7320

REFERENCES

1. Transportation Pooled Fund Study Details. <http://www.pooledfund.org/Details/Study/487>.
2. Urbanik, T., S. Beaird, D. Gettman, L. Head, D. Bullock, E. Smaglik, R. Campbell, and M. Ablett. *Traffic Signal State Transition Logic Using Enhanced Sensor Information*. Final Report, NCHRP 3-66, 2006.
3. Bonneson, J. A., A. Sharma, and D. Bullock. *Measuring the Performance of Automobile Traffic on Urban Streets*. Final Report, NCHRP 3-79, 2007.
4. Day, C. M. and D. Bullock. Arterial Performance Measures, Volume 1: Performance Based Management of Arterial Traffic Signal Systems. Final Report, NCHRP 3-79A (Vol. 1), 2010.
5. Day, C. M., E. J. Smaglik, D. M. Bullock, and J. R. Sturdevant. *Real-Time Arterial Traffic Signal Performance Measures*. Publication FHWA/IN/JTRP-2008/09. Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, West Lafayette, Indiana, 2008. doi: [10.5703/1288284313439](https://doi.org/10.5703/1288284313439).
6. Day, C. M., T. M. Brennan, H. Premachandra, A. M. Hainen, S. M. Remias, J. R. Sturdevant, G. Richards, J. S. Wasson, and D. M. Bullock. *Quantifying Benefits of Traffic Signal Retiming*. Publication FHWA/IN/JTRP-2010/22. Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, West Lafayette, Indiana, 2010. doi: [10.5703/1288284314250](https://doi.org/10.5703/1288284314250).
7. Day, C. M., T. M. Brennan, J. M. Ernst, J. Sturdevant, and D. M. Bullock. *Procurement Procedures and Specifications for Performance Measure Capable Traffic Infrastructure Data Collection Systems*. Publication FHWA/IN/JTRP-2011/18. Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, West Lafayette, Indiana, 2011. doi: [10.5703/1288284314642](https://doi.org/10.5703/1288284314642).
8. Liu, H. X., W. Ma, X. Wu, and H. Hu. Development of a Real-Time Arterial Performance Monitoring System Using Traffic Data Available from Existing Signal Systems. Report MN/RC 2009-01, University of Minnesota, 2008.
9. Institute of Transportation Engineers. *Advanced Transportation Controller*. Available online at <http://www.ite.org/standards/atc/>.
10. Signal Control Program Environment website. <https://sites.google.com/site/atiscopes/>.
11. Luk, J. The Interface Between Traffic Control Software and Control Module Software in the AATSC. Publication AP-T223-12, Austroads Ltd., Sydney, Australia, 2012.
12. Tyack, F. G. Street Traffic Signals, With Particular Reference to Vehicle Actuation. *IEE Journal*, Vol. 82, 1938, pp. 125–154.
13. Sims, A. G., and K. W. Dobinson. The Sydney Coordinated Adaptive Traffic System Philosophy and Benefits. *IEEE Transactions on Vehicular Technology*, Vol. 29, 1980, pp. 130–137.
14. Stevanovic, A. *Adaptive Traffic Control Systems: Domestic and Foreign State of Practice*. NCHRP Synthesis 403. Transportation Research Board of the National Academies, Washington, D.C., 2010. <http://www.trb.org/Publications/Blurbs/163323.aspx>.
15. Bullock, D. M., C. M. Day, T. M. Brennan, J. R. Sturdevant, and J. S. Wasson. Architecture for Active Management of Geographically Distributed Signal Systems. *ITE Journal*, Vol. 81, No. 5, 2011, pp. 20–24. Accession Number: 01352868
16. Koukoumidis, E., L.-S. Peh, and M. Martonosi. “SignalGuru: Leveraging Mobile Phones for Collaborative Traffic Signal Schedule Advisory.” *MobiSys 2011*, Bethesda, Maryland, USA, 2011.
17. Green Driver Mobile Phone Application. Available online at <http://imagreendriver.com>.
18. Remias, S. M., A. M. Hainen, G. M. Grimmer, A. D. Davis, C. M. Day, T. M. Brennan and D. M. Bullock. Leveraging High Resolution Signalized Intersection Data to Characterize Discharge Headway Distributions and Saturation Flow Rate Reliability. Transportation Research Board Annual Meeting, Paper No. 12-2122, 2011.
19. Smaglik, E., S. Vanjari, V. Totten, E. Rusli, M. Ndoye, D. M. Bullock, and J. V. Krogmeier. Performance of Modern Stop Bar Loop Count Detectors over Various Traffic Regimes. Transportation Research Board Annual Meeting, Paper 07-1090, 2007.
20. Using Existing Loops at Signalized Intersections for Traffic Counts: An ITE Informational Report. *ITE Journal*, Vol. 78, No. 2, Feb. 2008.
21. Smaglik, E. J., Z. Davis, R. C. Steele, W. Nau, and C. A. Roberts. Supplementing Signalized Intersection Infrastructure to Provide Automated Performance Measures with Existing Video Detection Equipment. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2192, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 77–88. doi: [10.3141/2192-07](https://doi.org/10.3141/2192-07).
22. Smaglik, E. J., A. Sharma, D. M. Bullock, J. R. Sturdevant, and G. Duncan. Event-Based Data Collection for Generating Actuated Controller Performance Measures.” In *Transportation Research Record No. 2035*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 97–106. doi: [10.3141/2035-11](https://doi.org/10.3141/2035-11).
23. Sturdevant, J. R., T. Overman, E. Raamot, R. Deer, D. Miller, D. M. Bullock, C. M. Day, T. M. Brennan, H. Li, A. Hainen, and S. M. Remias. *Indiana Traffic Signal Hi Resolution Data Logger Enumerations*. Indiana Department of Transportation and Purdue University, West Lafayette, Indiana, 2012. doi: [10.5703/1288284315018](https://doi.org/10.5703/1288284315018).
24. Sunkari, S. R., H. A. Charara, and P. Songchitruska. A Portable Toolbox to Monitor and Evaluate Signal Operations. Transportation Research Board Annual Meeting, Paper 12-2900, 2012.
25. Barkley, T., R. Hranac, K. Fuentes, and P. Law. Heuristic Approach for Estimating Arterial Signal Phases and Progression Quality from Vehicle Arrival Data. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2259, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 48–58. doi: [10.3141/2259-05](https://doi.org/10.3141/2259-05).
26. TrafInfo Communications, Inc. website. <http://www.trafinfo.com>.
27. Cesme, B. and P. G. Furth. Self-Organizing Control Logic for Oversaturated Arterials. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2356, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 92–99. doi: [10.3141/2356-11](https://doi.org/10.3141/2356-11).
28. Koonce, P., L. Rodegerdts, K. Lee, S. Quayle, S. Beaird, C. Braud, J. Bonneson, P. Tarnoff, and T. Urbanik. *Traffic Signal Timing Manual*. Report FHWA-HOP-08-024. Federal Highway Administration, Washington, D.C., 2008.
29. Furth, P.G., T. H. J. Muller, M. Salomons, T. A. Bertulis, and P. J. Koonce. Barrier-Free Ring Structures and

- Pedestrian Overlaps in Signalized Intersection Control. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2311, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 132–141. doi: [10.3141/2311-13](https://doi.org/10.3141/2311-13).
30. Day, C. M. Performance-Based Management of Arterial Traffic Signal Systems. PhD Thesis, Purdue University, 2010.
 31. Brennan, T. M., B. D. Griggs, G. Grimmer, A. M. Hainen, C. M. Day, J. R. Sturdevant, and D. M. Bullock. Defining Design Space for Parameters of Traffic Signal Timing: Empirical Approach. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2311, Transportation Research Board of the National Academies, Washington, D.C., No. 2311, 2012, pp. 85–98. doi: [10.3141/2311-08](https://doi.org/10.3141/2311-08).
 32. Wasson, J., M. Abbas, D. Bullock, A. Rhodes, and C. K. Zhu. *Reconciled Platoon Accommodations at Traffic Signals*. Publication FHWA/IN/JTRP-99/1. Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, West Lafayette, Indiana, 1999. doi: [10.5703/1288284313301](https://doi.org/10.5703/1288284313301).
 33. Jiang, Y., S. Li, and D. E. Shamo. A Platoon-Based Traffic Signal Timing Algorithm for Major-Minor Intersection Types. *Transportation Research Part B*, Vol. 40, 2006, pp. 543–562. doi: [10.1016/j.trb.2005.07.003](https://doi.org/10.1016/j.trb.2005.07.003).
 34. Gershenson, C. Self-Organizing Traffic Lights. *Complex Systems*, Vol. 16, 2005, pp. 29–53.
 35. Day, C. M., E. J. Smaglik, D. M. Bullock, and J. R. Sturdevant. Quantitative Evaluation of Fully Actuated Versus Non-Actuated Coordinated Phases. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2080, Transportation Research Board of the National Academies, Washington, D.C., 008, pp. 8–21. doi: [10.3141/2080-02](https://doi.org/10.3141/2080-02).
 36. Day, C. M. and D. M. Bullock. Design Guidelines and Conditions that Warrant Deploying Fully Actuated Coordination. Transportation Research Board Annual Meeting, Paper No. 14-0208.
 37. National Transportation Communications for ITS Protocol: Object Definitions for Actuated Traffic Signal Controller Units—Version 02. American Association of State Highway and Transportation Officials, Washington, D.C., 2005.
 38. Roupail, N., A. Tarko, and J. Li. Traffic Flow at Signalized Intersections. In *Traffic Flow Theory Monograph*, ed. N. Gartner, C. J. Messer, and A. K. Rath, Federal Highway Administration, Washington, D.C., 2001, <http://www.fhwa.dot.gov/publications/research/operations/tft/>.
 39. Webster, F.V. *Traffic Signal Settings*, Technical Paper No. 39, Road Research Laboratory, Her Majesty's Stationery Office, London, 1958.
 40. *Highway Capacity Manual*. Transportation Research Board of the National Academies, Washington, D.C., 2010.
 41. Perez-Cartagena, R. I. and A. Tarko. *Predicting Traffic Conditions at Indiana Signalized Intersections*. Report FHWA/IN/JTRP-2004/29, Joint Transportation Research Program, Purdue University, 2004. doi: [10.5703/1288284313294](https://doi.org/10.5703/1288284313294).
 42. C. M. Day, J. M. Ernst, T. M. Brennan, C. Chou, A. M. Hainen, S. M. Remias, A. Nichols, B. D. Griggs, and D. M. Bullock. Performance Measures for Adaptive Signal Control: Case Study of Central System-in-the-Loop Simulation. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2311, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 1–15. doi: [10.3141/2311-01](https://doi.org/10.3141/2311-01).
 43. Remias, S. M., A. M. Hainen, G. M. Grimmer, A. D. Davis, C. M. Day, T. M. Brennan, and D. M. Bullock. Leveraging High Resolution Signalized Intersection Data to Characterize Discharge Headway Distributions and Saturation Flow Rate Reliability. Transportation Research Board Annual Meeting, Paper 12-2122, 2012.
 44. Day, C. M., J. R. Sturdevant, H. Li, A. M. Hainen, S. M. Remias, and D. M. Bullock. Revisiting the Cycle Length—Lost Time Question With Critical Lane Analysis. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2355, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 1–9. doi: [10.3141/2355-01](https://doi.org/10.3141/2355-01).
 45. Li, H., A. M. Hainen, C. M. Day, G. Grimmer, J. R. Sturdevant, and D. M. Bullock. Longitudinal Performance Measures for Assessing Agency-Wide Signal Management Objectives. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2355, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 20–30. doi: [10.3141/2355-03](https://doi.org/10.3141/2355-03).
 46. Luyanda, F., D. Gettman, L. Head, S. Shelby, D. Bullock, and P. Mirchandani. ACS-Lite Algorithmic Architecture: Applying Adaptive Control System Technology to Closed-Loop Systems. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1856, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 175–184. doi: [10.3141/1856-19](https://doi.org/10.3141/1856-19).
 47. Smaglik, E. J., D. Gettman, D. M. Bullock, C. M. Day, and H. Premachandra. Comparison of Alternative Real-Time Performance Measures for Measuring Signal Phase Utilization and Identifying Oversaturation. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2259, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 123–131. doi: [10.3141/2259-11](https://doi.org/10.3141/2259-11).
 48. Freije, R. S., A. M. Hainen, A. Stevens, H. Li, W. B. Smith, C. M. Day, J. R. Sturdevant, and D. M. Bullock. Graphical Performance Measures for Practitioners to Triage Split Failure Trouble Calls. Transportation Research Board Annual Meeting, Paper 14-0301, 2014.
 49. Crommelin, R. W. Employing Intersection Capacity Utilization Values to Estimate Overall Level of Service. *Traffic Engineering*, 44, 1974, pp. 11–14.
 50. Day, C. M., D. M. Bullock, and J. R. Sturdevant. Cycle Length Performance Measures: Revisiting and Extending Fundamentals. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2128, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 48–57. doi: [10.3141/2128-05](https://doi.org/10.3141/2128-05).
 51. Quiroga, C. A. and D. Bullock. Measuring Control Delay at Signalized Intersections. *Journal of Transportation Engineering*, Vol. 125, 1999, pp. 271–280. doi: [10.1061/\(ASCE\)0733-947X\(1999\)125:4\(271\)](https://doi.org/10.1061/(ASCE)0733-947X(1999)125:4(271)).
 52. Hillier, J. A. and R. Rothery. The Synchronization of Traffic Signals for Minimum Delay. *Transportation Science*, Vol. 1, 1967, pp. 81–94. doi: [10.1287/trsc.1.2.81](https://doi.org/10.1287/trsc.1.2.81).
 53. Sharma, A., D. M. Bullock, and J. A. Bonneson. Input-Output and Hybrid Techniques for Real-Time Prediction of Delay and Maximum Queue Length at Signalized Intersections. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2035, Transportation Research Board of the National

- Academies, Washington, D.C., 2007, pp. 69–80. doi: [10.3141/2035-08](https://doi.org/10.3141/2035-08).
54. Day, C. M. and D. M. Bullock. Application of High Resolution Traffic Signal Controller Data for Platoon Visualization and Optimization of Signal Offsets. Presented at the Mobil.TUM International Scientific Conference on Mobility and Transport, Munich, Germany, 2009.
 55. Day, C. M., R. J. Haseman, H. Premachandra, T. M. Brennan, J. S. Wasson, J. R. Sturdevant, and D. M. Bullock. Evaluation of Arterial Signal Coordination: Methodologies for Visualizing High-Resolution Event Data and Measuring Travel Time. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2192*, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 37–49. doi: [10.3141/2192-04](https://doi.org/10.3141/2192-04).
 56. Brennan, T. M., C. M. Day, J. R. Sturdevant, and D. M. Bullock. Visual Education Tools To Illustrate Coordinated System Operation. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2259*, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 59–72. doi: [10.3141/2259-06](https://doi.org/10.3141/2259-06).
 57. Hainen, A. M., A. Stevens, C. M. Day, J. R. Sturdevant, and D. M. Bullock. High-Resolution Event-Based Data at Diamond Interchanges: Optimizing Ring Displacement. Transportation Research Board Annual Meeting, Paper 14-0298, 2014.
 58. Pacey, G. M. *Progress of a Bunch of Vehicles Released from a Traffic Signal*. Report RN/2665, Transport and Road Research Laboratory, Crowthorne, Berkshire, U.K., 1956.
 59. Robertson, D. I. *Transyt: A Traffic Network Study Tool*. Report LR 253, Road Research Laboratory, Crowthorne, Berkshire, Berkshire, U.K., 1969.
 60. Day, C. M. and D. M. Bullock. Computational Efficiency of Alternative Algorithms for Arterial Offset Optimization. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2259*, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 37–47. doi: [10.3141/2259-04](https://doi.org/10.3141/2259-04).
 61. Hunt, P. B., D. I. Robertson, R. D. Bretherton, and R. I. Winton. *SCOOT—A Traffic Responsive Method of Coordinating Signals*. Report No. LR 1014, Transport and Road Research Laboratory, Crowthorne, Berkshire, U.K., 1981.
 62. Abbas, M., D. Bullock, and L. Head. Real-Time Offset Transitioning Algorithm for Coordinating Traffic Signals. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1748*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 26–39. doi: [10.3141/1748-04](https://doi.org/10.3141/1748-04).
 63. Yin, Y., M. Li, and A. Skabardonis. Offline Offset Refiner for Coordinated Actuated Signal Control Systems. *Journal of Transportation Engineering*, Vol. 133, 2007, pp. 423–432.
 64. Day, C. M., T. M. Brennan, H. Premachandra, J. R. Sturdevant, and D. M. Bullock. Analysis of Peer Data on Intersections for Decisions About Coordination of Arterial Traffic Signal. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2259*, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 23–36. doi: [10.3141/2259-03](https://doi.org/10.3141/2259-03).
 65. Day, C. M., and D. M. Bullock. Calibration of Platoon Dispersion Model with High-Resolution Signal Event Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2311*, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 16–28. doi: [10.3141/2311-02](https://doi.org/10.3141/2311-02).
 66. Bonneson, J. A., M. P. Pratt, and M. A. Vanderhey. Predicting Arrival Flow Profiles and Platoon Dispersion for Urban Street Segments. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2173*, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 28–35. doi: [10.3141/2173-04](https://doi.org/10.3141/2173-04).
 67. Liu, H. X., X. Wu, W. Ma, and H. Hu. Real-Time Queue Length Estimation for Congested Signalized Intersections. *Transportation Research Part C*, Vol. 17, 2009, pp. 412–427. doi: [10.1016/j.trc.2009.02.003](https://doi.org/10.1016/j.trc.2009.02.003).
 68. Hubbard, S. M. L., D. M. Bullock, and C. M. Day. Integration of Real-Time Pedestrian Performance Measures into Existing Traffic Signal System Infrastructure. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2080*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 37–47. doi: [10.3141/2080-05](https://doi.org/10.3141/2080-05).
 69. Day, C. M., H. Premachandra, and D. M. Bullock. Characterizing the Impacts of Phasing, Environment, and Temporal Factors on Pedestrian Demand at Traffic Signals. Transportation Research Board Annual Meeting, Washington, D.C., Paper No. 11-0220.
 70. Brennan, T. M., C. M. Day, D. M. Bullock, and J. R. Sturdevant. Performance Measures for Railroad Preempted Intersections. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2128*, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 20–34. doi: [10.3141/2128-03](https://doi.org/10.3141/2128-03).
 71. Brennan, T. M., C. M. Day, J. R. Sturdevant, E. Raamot, and D. M. Bullock. Track Clearance Performance Measures for Railroad Preempted Intersections. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2192*, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 64–76. doi: [10.3141/2192-06](https://doi.org/10.3141/2192-06).
 72. Remias, S. M., A. M. Hainen, C. M. Day, T. M. Brennan, H. Li, E. M. Rivera, J. R. Sturdevant, S. E. Young, and D. M. Bullock. Performance Characterization of Arterial Traffic Flow with Probe Vehicle Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2380*, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 10–21. doi: [10.3141/2380-02](https://doi.org/10.3141/2380-02).
 73. Remias, S. M., T. Brennan, C. Day, A. Hainen, and D. Bullock. Characterizing Urban Mobility and Travel Time Reliability along Signalized Corridors using Probe Data. Presented at International Scientific Conference on Mobility and Transport, Munich, Germany, June 2013.
 74. Day, C. M., T. M. Brennan, A. M. Hainen, S. M. Remias, and D. M. Bullock. *Roadway System Assessment Using Bluetooth-Based Automatic Vehicle Identification Travel Time Data*. Purdue University Press, West Lafayette, Indiana, 2012. doi: [10.5703/1288284314988](https://doi.org/10.5703/1288284314988).
 75. Day, C. M., T. M. Brennan, A. M. Hainen, S. M. Remias, H. Premachandra, J. R. Sturdevant, G. Richards, J. S. Wasson, and D. M. Bullock. Reliability, Flexibility, and Environmental Impact of Alternative Objective Functions for Arterial Offset Optimization. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2259*, Transportation Research Board of the National

- Academies, Washington, D.C., 2011, pp. 8–22. doi: [10.3141/2259-02](https://doi.org/10.3141/2259-02).
76. Remias, S. M., A. M. Hainen, S. R. Mitkey, and D. M. Bullock. Probe Vehicle Re-Identification Data Accuracy Evaluation. ITS World Congress, Paper No. 3078, October 2011.
 77. Wasson, J. S., J. R. Sturdevant, and D. M. Bullock. Real-Time Travel Time Estimates Using Media Access Control Address Matching. *ITE Journal*, Vol. 78, No. 6, 2008, pp. 20–23. Accession Number 01108548.
 78. Brennan, T. M., J. M. Ernst, C. M. Day, D. M. Bullock, J. V. Krogmeier, and M. Martchouk. Influence of Vertical Sensor Placement on Data Collection Efficiency from Bluetooth MAC Address Collection. *Journal of Transportation Engineering*, Vol. 136, No. 12, 2010, pp. 1104–1109. doi: [10.1061/\(ASCE\)TE.1943-5436.0000178](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000178).
 79. Haseman, R. J., J. S. Wasson, and D. M. Bullock. Real-Time Measurement of Work Zone Travel Time Delay and Evaluation Metrics Using Bluetooth Probe Tracking. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2169*, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 40–53. doi: [10.3141/2169-05](https://doi.org/10.3141/2169-05).
 80. Wasson, J. S., G. W. Boruff, A. M. Hainen, S. M. Remias, E. A. Hulme, G. Farnsworth, and D. M. Bullock. Evaluation of Spatial and Temporal Speed Limit Compliance in Highway Workzones. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2258*, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 1–15. doi: [10.3141/2258-01](https://doi.org/10.3141/2258-01).
 81. Hainen, A. M., J. S. Wasson, S. M. L. Hubbard, S. M. Remias, G. D. Farnsworth, and D. M. Bullock. Estimating Route Choice and Travel Time Reliability with Field Observations of Bluetooth Probe Vehicles. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2256*, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 43–50. doi: [10.3141/2256-06](https://doi.org/10.3141/2256-06).
 82. Schrank, D., T. Lomax, and B. Eisele. *The 2011 Urban Mobility Report*. Texas Transportation Institute, Sept. 2011. <http://tti.tamu.edu/documents/mobility-report-2011.pdf>.
 83. Remias, S., T. Brennan, G. Grimmer, E. Cox, D. Horton, and D. Bullock. *2011 Indiana Interstate Mobility Report—Full Version*. Indiana Mobility Reports, 2012. doi: [10.5703/1288284314680](https://doi.org/10.5703/1288284314680).
 84. Remias, S., T. Brennan, C. Day, H. Summers, E. Cox, D. Horton, and D. Bullock. *2012 Indiana Interstate Mobility Report—Full Version*. Indiana Mobility Reports, 2013. doi: [10.5703/1288284315190](https://doi.org/10.5703/1288284315190).
 85. Liu, H. X., W. Ma, X. Wu, and H. Hu. Real-time Estimation of Arterial Travel Time Under Congested Conditions. *Transportmetrica*, Vol. 8, 2012, pp. 87–104.
 86. Liu, H. X. and W. T. Ma. A Virtual Vehicle Probe Model for Time-Dependent Travel Time Estimation on Signalized Arterials. *Transportation Research Part C*, Vol. 17, 2009, pp. 11–26. doi: [10.1016/j.trc.2008.05.002](https://doi.org/10.1016/j.trc.2008.05.002).
 87. Bhaskar, A., E. Chung, and A. G. Dumont. Fusing Loop Detector and Probe Vehicle Data to Estimate Travel Time Statistics on Signalized Urban Networks. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 26, 2011, pp. 433–450. doi: [10.1111/j.1467-8667.2010.00697.x](https://doi.org/10.1111/j.1467-8667.2010.00697.x).
 88. Ernst, J., C. M. Day, J. V. Krogmeier, and D. M. Bullock. Probe Data Sampling Guidelines for Characterizing Arterial Travel Time. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2315*, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 173–181. doi: [10.3141/2315-18](https://doi.org/10.3141/2315-18).

APPENDIX. REQUIRED DATA ELEMENTS FOR PERFORMANCE MEASURES

TABLE A.1
Performance measure data requirements.

Performance Measure	Description	Required Event Code IDs
Background Cycle Length (Figure 5.7)	Programmed cycle length as measured from time between successive yield points.	151. Yield Point
Effective Cycle Length (Figure 5.8)	Actual time that it takes to serve all phases in a cycle.	1. Phase Green
Green Time (Figure 5.11)	Actual green time displayed on a phase or overlap.	1. Phase Green 8. Phase Yellow
Capacity (Figure 5.13)	Green time scaled by saturation flow rate to derive the provided capacity.	32. FYA—Begin Permissive 33. FYA—End Permissive
<i>g/C</i> Ratio (Figure 5.15)	Ratio of green time to effective cycle length.	61. Overlap Green 63. Overlap Yellow
Vehicle Count (Figure 5.17)	Number of vehicles detected on a phase or overlap during a cycle.	1. Phase Green 8. Phase Yellow
Equivalent Hourly Volume (Figure 5.19)	Vehicle count scaled to vehicles per hour.	32. FYA—Begin Permissive 33. FYA—End Permissive
Volume-to-Capacity Ratio (Figure 5.21)	Equivalent hourly volume as a proportion of the provided capacity.	61. Overlap Green 63. Overlap Yellow 82. Detector On
Phase Termination (Figure 5.25)	Reason for phase termination in each cycle.	4. Phase Gap-Out 5. Phase Max-Out
Phase Termination Diagram (Figure 5.27)	Graphical plot of repeated phase force-offs.	6. Phase Force-Off
Green Occupancy Ratio (GOR) (Figure 5.28)	Proportion of green time that the detector is occupied.	1. Phase Green 8. Phase Yellow 32. FYA—Begin Permissive 33. FYA—End Permissive 61. Overlap Green 63. Overlap Yellow 81. Detector Off 82. Detector On
Red Occupancy Ratio of the First Five Seconds (ROR ₅)	Proportion of the first 5 seconds of red that the detector is occupied.	10. Phase Red Clearance 64. Overlap Red Clearance 81. Detector Off 82. Detector On
GOR/ROR Diagram (Figure 5.31)	Composite plot of GOR and ROR ₅ with phase termination. (Overlap plots might not include a “phase termination” option on the data.)	1. Phase Green 8. Phase Yellow 32. FYA—Begin Permissive 33. FYA—End Permissive 61. Overlap Green 63. Overlap Yellow 81. Detector Off 82. Detector On 4. Phase Gap-Out 5. Phase Max-Out 6. Phase Force-Off
Degree of Intersection Saturation (Figure 5.34)	Overall utilization of capacity provided by each phase in the critical path of the intersection.	1. Phase Green 8. Phase Yellow 81. Detector On
Percent on Green (Figure 6.5)	Proportion of vehicle arrivals taking place while the intersection is green.	1. Phase Green 8. Phase Yellow
Arrival Type (Figure 6.6)	A version of the percent on green that is divided by the green-to-cycle ratio and fitted to a qualitative 1–6 scale.	61. Overlap Green 63. Overlap Yellow 81. Detector On

TABLE A.1
(Continued)

Performance Measure	Description	Required Event Code IDs
Input-Output Delay (Figure 6.11)	An estimate of delay on an approach based on relationship between arrival profile and assumed departure profile.	1. Phase Green 8. Phase Yellow
Purdue Coordination Diagram (Figure 6.14)	A visualization of individual detector events relative to the status of the downstream phase or overlap.	61. Overlap Green 63. Overlap Yellow
Flow Profile (Figure 6.22)	Cyclic distributions of the probability of green and proportion of vehicle arrivals taking place during a cycle.	81. Detector On
Estimated Queue Length (Figure 6.31)	Estimated length of queue based on analysis of shockwaves and detector occupancy.	1. Phase Green 8. Phase Yellow 61. Overlap Green 63. Overlap Yellow 81. Detector Off 82. Detector On
Pedestrian Cycle (Figure 7.2)	Indication of whether a cycle included a pedestrian phase.	21. Phase Walk 67. Ped Overlap Walk
Pedestrian Actuation to Service Time (Figure 7.3)	Time between onset of a call for pedestrian service and beginning of pedestrian service.	21. Phase Walk 22. Phase Begin Ped Clear 23. Phase Solid Don't Walk 67. Ped Overlap Walk 68. Ped Overlap Begin Ped Clear 69. Ped Overlap Solid Don't Walk 90. Pedestrian Detector On
Pedestrian Conflicting Volume (Figure 7.4)	Volume on a movement that conflicts with a pedestrian phase.	(Same as Equivalent Hourly Volume)
Preemption Event Diagram (Figure 7.6)	Visualization of event durations relevant to preemption entry.	101. Preempt Advance Warning Input 102. Preempt Input On 103. Preempt Gate Down Input 104. Preempt Input Off 105. Preempt Entry Started 106. Preemption Begin Track Clearance 107. Preemption Begin Dwell Service
Preempt Duration (Figure 7.7)	Duration of preemption events.	101. Preempt Advance Warning Input 111. Preemption Begin Exit Interval
Priority Time to Green (Figure 7.8)	Time between onset of a call for transit priority and beginning of desired phase or overlap green.	112. TSP Check In 1. Phase Green 61. Overlap Green 8. Phase Yellow 63. Overlap Yellow
Detector Failure (Figure 8.5)	Histogram describing frequency of reported detector failures.	83. Detector Restored 84. Detector Fault—Other 85. Detector Fault—Watchdog 86. Detector Fault—Open Loop 87. Detector Fault—Shorted Loop 88. Detector Fault—Excessive Charge 91. Ped Detector Failed 92. Ped Detector Restored

ADDITIONAL NOTES:

To show boundaries between time-of-day periods, event 131, Coord Pattern Change could be used.

Effective cycle lengths could also potentially be measured using event 31, Barrier Termination. For controllers without a specific barrier event, the “barrier” could potentially be interpreted as a next phase transition from one compatibility group to another compatibility group.

Project Partners

This work was supported in part by the Pooled Fund Study TPF-5(258) led by the Indiana Department of Transportation (INDOT) and supported by the state transportation agencies of California, Georgia, Minnesota, Mississippi, New Hampshire, Pennsylvania, Texas, Utah, and Wisconsin, the Federal Highway Administration Office of Operations and Resource Center, and the Chicago Department of Transportation. This monograph summarizes performance measures developed by the Pooled Fund project and material developed previously by projects funded by the Joint Transportation Research Program, Traffax/USDOT (Project No. DTRT57-11-C-10009), and the National Cooperative Highway Research Program Project 3-79a. The authors also wish to thank the many agencies and vendors we have collaborated with during the development of data logging software and the infrastructure for harvesting the data.

The contents of this monograph 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.

Publication

This report was published in collaboration with the Joint Transportation Research Program and Purdue University. The full content of this technical report is available for download at <http://dx.doi.org/10.5703/1288284315333>. In addition, a print-on-demand bound version of this report can be purchased at that location.

Open Access and Collaboration with Purdue University

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 1500 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 2014, the JTRP collection has over 750,000 downloads, with some particularly popular reports having over 20,000 downloads.



Intersection of W. 5400 South and S. 2700 West, Salt Lake City, Utah.