1. Report No. FHWA/TX-04/0-4422-1	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle POTENTIAL MEASURES OF ASSESSING SIGNAL TIMING PERFORMANCE USING EXISTING TECHNOLOGIES		5. Report Date July 2004
		6. Performing Organization Code
7. Author(s) Kevin N. Balke and Curtis Herrick		8. Performing Organization Report No. Report 0-4422-1
9. Performing Organization Name and Address Texas Transportation Institute The Texas A&M University System College Station, Texas 77843-3135		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. Project No. 0-4422
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Implementation Office P. O. Box 5080 Austin, Texas 78763-5080		13. Type of Report and Period Covered Technical Report: September 2002-August 2003
		14. Sponsoring Agency Code

15. Supplementary Notes
Project performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration.
Project Title: *Measuring* Performance of Traffic Signal Systems Using Existing Detector Technology

16. Abstract
This report documents the first year of a two-year research project to develop potential measures that might be used to assess the performance of traffic signal timing at isolated intersections. This report summarizes the findings of a series of interviews we conducted to determine what measures the Texas Department of Transportation (TxDOT) is currently using to assess the performance of their traffic signals and how data for these performance measures are collected. It also contains the results of an examination of existing controller and detection technologies to determine what capabilities they have to produce performance measures. The report presents several proposed measures that can be used to assess the performance of signal timing and the procedures for computing these performance measures using existing detection technologies.

17. Key Words Performance Measures, Traffic Signals, Signal Timing	18. Distribution Statement No restrictions. This document is available to the public through NTIS: National Technical Information Service Springfield, Virginia 22161 http://www.ntis.gov		
19. Security Classif.(of this report) Unclassified	20. Security Classif.(of this page) Unclassified	21. No. of Pages 44	22. Price

Form DOT F 1700.7 (8-72)   Reproduction of completed page authorized

POTENTIAL MEASURES OF ASSESSING SIGNAL TIMING PERFORMANCE USING EXISTING TECHNOLOGIES

by

Kevin N. Balke, Ph.D., P.E.
Center Director, TransLink® Research Center
Texas Transportation Institute

and

Curtis Herrick, P.E.
Associate Research Engineer
Texas Transportation Institute

July 2004

TEXAS TRANSPORTATION INSTITUTE
The Texas A&M University System
College Station, Texas   77843-3135

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of the Federal Highway Administration (FHWA) or the Texas Department of Transportation (TxDOT). This report does not constitute a standard, specification, or regulation. The engineer in charge was Kevin N. Balke, P.E. (Texas, #66529).

ACKNOWLEDGMENTS

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER 1:
INTRODUCTION

INTRODUCTION

FHWA defines performance measurement (or monitoring) as the "use of statistical evidence to determine progress toward specific defined organizational objectives." Performance measures can be actual "hard and fast" measured parameters (such as pavement surface smoothness or travel times) or they may be measures of customer satisfaction (such as the perceived ride quality or on-time arrivals). Regardless of how they are measured, performance measures can be used to provide "feedback" about how well their system is performing, both from a user's and an operator's perspective (*1*).

With the enactment of the Intermodal Surface Transportation Efficiency Act (ISTEA) legislation in 1991, FHWA began to place a greater emphasis on using performance measures and performance monitoring. A recent National Cooperative Highway Research Program (NCHRP) study showed that while a wide range of possible applications exists for performance measures, they are primarily used for the following purposes:

- responses to legislative mandates,
- planning processes, including budget and funding allocations,
- quality initiatives,
- congestion management systems and evaluation,
- ITS operations and evaluations,
- safety management systems, and
- permit processes for commercial driveways.

In the past, much of the research related to performance measures has focused on infrastructure management and freeway operations. Table 1 shows the performance measures used by many agencies to evaluate the performance of highway sections and systems. While these performance measures are geared more toward freeways and highway sections, Table 1 shows the breadth and diversity of measures that agencies are using to assess the performance of their systems.

Table 1. Measures Commonly Used by Transportation Agencies to Assess Performance of Highway Segments and Systems

Performance Measure	Typical Definition
• Level of Service (LOS)	• Qualitative measurement of highway point, segment of system using A(best) to F(worst) based on measures of effectiveness
• Traffic Volume	• Actual average daily traffic, peak-hour traffic or peak-period traffic
• Vehicle-Miles Traveled	• Volume times length
• Travel Time	• Distance divided by speed
• Speed	• Distance divided by travel time
• Incidents	• Traffic interruption caused by a crash or other unscheduled event
• Duration of Congestion	• Period of congestion
• Percent of System Congested	• Percent of miles congested (usually defined based on LOS E or F)
• Vehicle Occupancy	• Persons per vehicle
• Percent of Travel Congested	• Percent of vehicle-miles or person-miles traveled
• Delay Caused by Incidents	• Increase in travel time caused by an incident
• Density	• Vehicles per lane per period
• Rail Crossing Incidents	• Traffic crashes that occur at highway-rail grade crossings
• Recurring Delay	• Travel time increases from congestion; this measure does not consider incidents
• Travel Costs	• Value of driver's time during a trip and any expenses incurred during the trip (vehicle ownership and operating expenses or tolls or tariffs)
• Weather-Related Traffic Incidents	• Traffic interruption caused by inclement weather
• Response Time to Incidents	• Period required for an incident to be identified, verified, and for an appropriate action to alleviate the interruption to traffic to arrive at the scene
• Commercial Vehicle Safety Violations	• Number of violations issued by law enforcement based on vehicle weight, size, or safety
• Evacuation Clearance Time	• Reaction and travel time for evacuees to leave an area at risk
• Response Time to Weather-Related Incidents	• Period required for an incident to be identified, verified, and for an appropriate action to alleviate the interruption to traffic to arrive at the scene
• Security for Highway and Transit	• Number of violations issued by law enforcement for acts of violence against travelers
	• Dollars generated by tolls
• Toll Revenue	• Several definitions are used that include (1) variability of travel times; (2) percent of travelers who arrive at their destination within an acceptable time, and (3) range of travel times
• Travel Time Reliability	
Source: NCHRP Report 311(2)	

OBJECTIVES

The overall goal of this project is to examine current and innovative methods of collecting measures that TxDOT can use to assess the performance of their traffic signals. The project is a two-year project, with the first year focusing on 1) analyzing the capabilities of existing technology, and 2) assessing TxDOT's needs for measures related to the performance of traffic signals. The objectives of the first year of this project were as follows:

- Through interviews, identify how TxDOT engineers and traffic signal technicians assess the performance of their traffic signals in the field.
- Assess the capabilities of the existing detection and traffic signal controller technology to provide these measures
- If necessary, propose new and innovative measures for evaluating the performance of traffic signals.

ORGANIZATION OF REPORT

This report summarizes the results of the first year of a two-year project on the use of measures to assess the performance of traffic signals. Chapter 2 of this report summarizes the findings of a series of interviews we conducted to determine what measures TxDOT is currently using to measure the performance of its traffic signals and how data for developing these performance measures are collected. Chapter 3 contains the results of an examination of existing controller and detection technologies to determine what capabilities they have to produce performance measures. In Chapter 4 we discuss several proposed performance measures and how they can be computed using existing detection technologies.

CHAPTER 2:
SUMMARY OF FINDINGS FROM SITE VISITS

INTRODUCTION

Site visits were conducted in three TxDOT districts: Houston, Lufkin, and Pharr. The purpose of these site visits was to discuss with TxDOT traffic engineers and signal technicians the process and measures they use to assess how well a traffic signal is operating in their district. This chapter provides a summary of the findings of these sites visits.

DATA USED TO DEVELOP TRAFFIC SIGNAL TIMING PLANS AND ASSESS PERFORMANCE

One of the first issues discussed in the site visits was what type of data TxDOT used to develop and evaluate traffic signal timing plans. All of the districts visited indicated that volume data (and more specifically, turning movement volume counts) were used to develop and evaluate signal timing plans. Other studies cited as being used include the following:

- delay studies
- gap studies, and
- travel time studies.

All of the districts indicated that they do not have enough resources available to them to always collect the data they need to develop good timing plans or assess performance of existing timing plans. In some cases, traffic signal timing decisions are made using data that are 3-, 5-, and sometimes 10-years old. In other cases, data from other sources, such as traffic impact studies or obtained from consultants, are used to develop signal timing plans. All three districts indicated that they commonly use organizations from outside of TxDOT to collect traffic data, and this is primarily done on an as-needed basis.

Furthermore, all of the districts indicated that they do not have a regular program to evaluate the effectiveness of their traffic signal timings at their intersections. Instead, studies to assess the performance of their signals timings were generally initiated in response to citizen complaints. Again, all three districts indicated that was due not from a lack of desire or perceived benefit from such a program, but a lack of adequate staffing and resources to execute and maintain a program.

USE OF OPTIMIZATION AND SIMULATION MODELS

Representatives from each district were asked about the type of simulation and optimization models they used to assist them in establishing initial signal timing plans and to assess performance. This was important to this project because it allowed the researchers insight into the types of performance measures engineers and technicians are used to working with in their local areas.

All of the districts indicated that they are familiar with and used at least one of the models in the PASSER family (i.e., PASSER II, PASSER III, or PASSER IV, depending upon the situation) as well as SYNCHRO$^®$ to optimize traffic signal timings. The PASSER models were used primarily to develop signal timing plans for networks and diamond interchanges while the SYNCHRO$^®$ model was used to develop timings for isolated intersections and arterials. This is important because these models use different criteria for optimizing performance of traffic signals. For example, the PASSER II model attempts to optimize signal performance based on progression bandwidth and attainability of progression. PASSER III attempts to balance queue storage at diamond interchanges. SYNCHRO$^®$, on the other hand, attempts to minimize delay at intersections as one of its primary optimization parameters. Where different engineers and technicians use optimization models provides insight into the goals and objectives that they are trying to achieve when developing signal timing plans for a particular intersection.

None of the districts indicated that they regularly use simulation as a way of assessing performance. In the rare occasion that simulation is needed, they generally rely on a consultant to perform the analysis for them.

PARAMETERS USED TO EVALUATE EFFECTIVENESS OF TIMING PLANS

District surveys asked personnel how they evaluated the effectiveness of the timing plans they were using in their districts. Table 2 lists all the measures that were used to gauge the effectiveness of the operations at intersections.

Table 2. Common Measures Cited by TxDOT District Personnel for Evaluating the Effectiveness of Signal Timing Plans.

• Citizen complaints	• Cross-street delay
• Percent of cycle used	• Green utilization
• Occupancy	• Number of vehicles serviced during green interval
• Number of stops	
• Cycle failures	• Number of vehicles remaining after green interval
• Control delay	
• Queue length	• Queue clearance
• Traffic demand	• Number of gap outs
• Volume	• Average duration of green interval
• Speed	• Departure headways

All of the districts indicated that citizen complaints were their primary means of learning about intersections where potential timing problems might exist. Citizen complaints serve as an indicator that something needs to be done at an intersection. Unfortunately, citizen complaints, by themselves, often describe only the symptoms of the problem and not the source of the problem. For example, a citizen may complain about the delay in being serviced at an intersection. In most cases, it is impossible to determine if the source of the problem is poor signal timing or malfunctioning detection equipment.

At least in one district, clearing the queues on the cross-street seems to be the primary signal timing objective in under-saturated conditions. This philosophy often results in "longer than normal" maximum green times, but because during most cycles the cross-street approaches "gap out," the signal "returns early" to the main street approaches, which in turn provides more green time to the main-street approach under light-volume conditions.

Several districts indicated that cycle failures may not necessarily be a good performance measure because of pedestrian activities and the "resync" issues as the controller returns to coordination.

Queue length was the most frequently cited measures for evaluating the effectiveness of traffic signal timing in oversaturated conditions. In oversaturated conditions, the primary signal timing objective seems to change from one of delay minimization to one of balancing queue

length and queue growth. One district specifically cited that in oversaturated conditions, they use a philosophy of "no one movement suffers more than the others." To achieve this objective, they use a combination of queue length and vehicle delay to strike a balance between the movements at an intersection. One district indicated that what is really needed to measure the effectiveness of signal timing strategy in oversaturated conditions is not queue length, but how many vehicles were serviced during the green interval and how many vehicles remained in the queue after the green interval terminated.

REQUIREMENTS OF AN AUTOMATED PERFORMANCE MEASURE SYSTEM

A final issue discussed as part of the site visits was what type of requirements the district had of an automated system for collecting traffic signal timing performance measures. Listed below are some of the requirements identified by district personnel:

- The system must produce simplified results that can be easily presented to others.
- The system must be easy to setup and data must be easy to retrieve.
- The system should be designed to be operated by lower-skilled technicians. The system should have effective and responsive help screens.
- If it is to be incorporated into traffic signal system vendor software, it needs to be widely available on all controllers and it needs to be implemented in a similar manner and produce similar output. Consistency is essential.
- The system needs to be able to store data for future retrieval. It should never have to store more than one month of data at the intersection level. The system should have upload capabilities.
- The user should be able to define the time slice (or duration) over which the data is collected and reported. The user should be able to select where the data should be collected or displayed on a per-period, per-hour, or per-cycle basis.
- The performance measures should be relatively easy to understand, not only by technicians but also by the motoring public.

CHAPTER 3:
ASSESSMENT OF AVAILABLE TECHNOLOGIES

INTRODUCTION

Before embarking on developing a new system to collect performance measures, the researchers first conducted an assessment of the capabilities of the existing traffic signal and detection technologies to collect and report performance measures. This chapter summarizes some of the capabilities of that equipment as it relates to measuring the performance of traffic signal control.

EAGLE EPAC TRAFFIC SIGNAL CONTROLLER

Of the two traffic signal vendors prequalified by TxDOT, only the Eagle EPAC300 Actuated Controller has a built-in process for generating a limited amount of performance measure data (other than volume and occupancy reports from system detectors). As part of its normal logging and reporting capability, the Eagle EPAC300 Actuated Traffic Signal Controller Unit automatically produces two reports that could potentially be used to assess the performance of signal operations at an isolated intersection: the Measures Of Effectiveness (MOE) Report and the Cycle MOE Report *(3)*. The MOE Report is a report produced by the controller that is intended to be used to evaluate the effectiveness of coordination (namely green split). The MOE Report contains the following measures for each phase (Phase 1-8) for up to 24 coordination plans [i.e., the different Dial/Split/Offset (D/S/O) combination used at an intersection]:

- volume,
- stops,
- delays, and
- utilization.

Eagle defines "volume" as the "average number of actuations during the sequence cycle for the duration of the pattern." It is computed by averaging the total number of vehicle actuations that occur each cycle by the total number of cycles that occur while the coordination plan is in effect. For example, if a particular coordination plan (i.e., D/S/O = 1/1/1) was active for three cycles and, during those three cycles, the controller detected 25, 30, and 20 vehicles, respectively, the computed volume would be 25 vehicles.

Eagle defines "stops" as "the average number of vehicles that must stop at an intersection during the cycle duration of the pattern." It is computed by averaging the number of actuations that the controller receives during the red interval each cycle by the total number of cycles that occur while a particular plan is active. In other words, if during three cycles the controller detected 15, 5, and 10 vehicles stopping on an approach, then the computed number of stops would be equal to 10 vehicles.

The EPAC300 also computes the delay for each phase in a timing plan. Delay is defined as the average time, in seconds, that vehicles are stopped during the sequence cycle for the duration of the pattern. It is computed by averaging the accumulated wait time (i.e., the number of vehicles "waiting" multiplied by time) for each phase per sequence cycle.

Utilization is the final measure contained in the MOE report. Eagle defines utilization as the average number of seconds that the green interval lasts for each phase. It is computed by measuring duration of the green interval for each phase each cycle and then dividing it by the total number of cycles that occur while a coordination plan is active. For example, if for over three cycles the green interval lasted for 12, 14, and 16 seconds, respectively, then the utilization would be reported as 14.

In addition to the MOE Report, the Eagle EPAC300 also produces a report called the Cycle MOE Report. This report shows how much each phase varied from its programmed split time each cycle. For example, if the split for a given phase is programmed to be 20 seconds but the phase was actually active for 25 seconds, a value of 5 would appear in the cycle report under that particular phase. Positive entries in the report imply that actual time that a phase was active was longer than the programmed split, while negative values imply that the phase actually ran shorter than the programmed split. Every time a new event associated with coordination occurs, a new entry occurs in this report. Therefore, a traffic signal technician will find entries in this report when the following events occur:

- at the beginning of a new cycle (local cycle time equal to zero),
- when the controller transitions to a new coordinated signal timing plan, and
- when controller transitions from coordinated operation to some other mode of operations (including PREEMPT and FREE).

While these reports produce performance measures that districts could potentially use to assess the effectiveness of signal timings, the researchers have identified several possible limitations that might restrict the application of these reports:

- Many of the TxDOT traffic signal timing engineers were not aware of the capabilities of the controller system to produce these performance measures.
- Both of the MOE Report and the Cycle Report are available only when the controller in running in coordination (i.e., the COORD mode). When the controller is running in the FREE mode (i.e., as an actuated, isolated intersection), these performance measures are NOT computed.
- The manner in which these performance measures are computed seems best suited for one-lane approaches only. TxDOT's current detectorization scheme involves using multiple detectors in multiple lanes to make calls to the controller on an approach basis. What is truly needed is to improve the level of accuracy of these performance measures in detectorization scheme that keeps track of vehicle arrivals on a lane-by-lane basis.

A simulation study is currently underway to assess the accuracy of the performance measures in a multi-loop, multi-lane intersection.

AUTOSCOPE SOLO® SYSTEM

At many new intersections, TxDOT is installing a video imaging vehicle detection system (VIVDS) instead of loop detectors. VIVDS uses video imaging technology to emulate and enhance the detection capabilities of most standard loop detectors.   With VIVDS, detection zones are established in software (as opposed to physically cutting them into the pavement) on approaches to intersections. These detection zones are then used to provide the traffic signal controller with calls for service.

One particular vendor of VIVDS that TxDOT commonly uses by is the Autoscope Solo® system (3). Using Autoscope®, a traffic engineer or signal technician can provide up to 150 different combinations of detection zones and, depending upon the type of traffic signal (TS) controller cabinet (TS-1 or TS-2) and intra-cabinet communications architecture (Rack Card, Mini-Hub, Hub, etc.), anywhere from 8 to 64 different inputs into a traffic signal controller.

These outputs can be used to call various vehicular and pedestrian phases as well as provide inputs into typical system detectors.

Three basic types of detectors are used at intersection applications: count detectors, presence detectors, and speed detectors. Count detectors are specifically designed to measure the volume of traffic on an approach by counting the number of vehicles that pass under a detection zone. Volume is computed by summing all the vehicle detections that occur over a user-specified interval. Count detectors are generally placed perpendicular to the stop line of the intersection and can be placed anywhere in the field of view of the camera.

Presence detectors can be used to identify whether or not a vehicle is present in the field of view. Presence detectors can be established to look for traffic moving in any direction or for vehicles traveling in one direction only.   In most intersection applications, presence detectors are typically installed parallel to the flow of traffic.   Presence detectors can also be set up to detect vehicles that are stopped. Autoscope® defines a stopped vehicle as one that has remained in the detection zone for 3 seconds or longer.

Speed detectors can be installed to measure the speed of individual vehicles (in miles per hour or kilometers per hour), the length of vehicles, and the classification of vehicles. The software computes speed by measuring the time it takes for an individual vehicle to traverse a detection zone of a known length (i.e., the time of the front of the vehicle to travel from the front edge of the detection zone to the rear edge of the detection zone). Vehicle length is computed by multiplying the vehicle speed by the time differential between the passage of the front of the vehicle and the rear of the vehicle at some point in the detection zone (either the leading edge or the trailing edge of the detection zone).   Vehicle length can then be used to classify vehicles into five user-defined categories.

By combining various combinations of count, presence, and speed detectors into detector stations, traffic data can be gathered over a user-specified interval. Using detector stations, the Autoscope® system can automatically compute the following types of traffic flow information:

- *Average Flow Rate* – defined as the equivalent hourly rate at which vehicles pass under the count detector, in vehicles per hour;
- *Total Volume Count* – defined as the total number of vehicles that passed under the detector during the polling interval;

- *Arithmetic Mean Speed* – defined as the average speed of all the vehicles passing under a speed detector during the polling interval;

- *Vehicle Class Count* – defined as the number of vehicles in each of the five vehicle classifications;

- *Average Time Headway* – defined as the average passage time between two successive vehicles;

- *Average Time Occupancy* – defined as the percentage of time that vehicles are present at a detector during the polling interval;

- *Level of Service* – the listing of the level of service (A, B, C, D, E, or F) based upon either speed or approach capacity;

- *Space Mean Speed* – obtained by dividing the total distance traveled by tow or more vehicles on a section of highway by the total time required by those vehicles to travel that distance;

- *Space Occupancy* – defined as the amount of a given section of highway occupied by vehicles of a given length at an instant in time, expressed as a percentage; and

- *Density* – calculated as the number of vehicles divided by the distance between two points on a highway at an instant in time.

The parameters can be computed over a user-defined time interval (for example, 1 minute, 10 minutes or 1 hour) or can be computed on a cycle-by-cycle basis. The user can also select which controller phase input to use to accumulate the performance measures. Generally, detection zones are established on a per-lane basis.

One interesting feature of the Autoscope® system is that it can be tied to outputs of the traffic signal control so that it will activate different detectors during different portions of the phase or cycle. For example, the Autoscope® system allows the engineer/technician to activate a detector depending upon whether the signal on an approach is displaying a green or red indication.

ORACLE /2 INDUCTIVE LOOP SYSTEM

Most of the inductive loop detectors commonly used by TxDOT are capable of producing volume and occupancy counts. Volume counts can be produced by accumulating pulses (or detector calls) as vehicles pass over the loop. Occupancy can be defined as the percentage of time that a vehicle occupies a detection zone.

Generally, loop detectors are operated in one of two modes: presence mode or pulse mode. In presence mode, the output of the detector remains energized (or "ON") as long as a vehicle is present in the detection zone. Once the vehicle leaves the detection zone, the output of the detector is de-energized (or "OFF"). In the pulse mode, the detector output is "pulsed" active for 125 ms when a vehicle enters the detection area. Unlike a detector that is operated in the presence mode, in the pulse mode, the output of the detector is pulsed again with each sequent vehicle arrival, regardless of how many vehicles are already occupying the detection area. Regardless of the mode of operation, volume is determined by counting the number of "ON-OFF" sequences while occupancy is determined by measuring percentage of time the detector is in the "ON" state.

The ORACLE /2 is a programmable, dual channel inductive loop vehicle detector that is designed and manufactured by Eberle Design, Inc. *(4)*. The ORACLE /2 also has a feature called "AccurateCount" which enables vehicles to be counted and totals displayed on the units LCD display. If the detector is set to operate in the AccurateCount mode, it produces a secondary output, in addition to the primary CALL output, every time a vehicle enters the loop detection zone. Each vehicle entering the loop will cause an output pulse of 125 ms $\pm$ 25 ms, regardless of the size of the loop. This feature allows each vehicle to enter the detection zone even if a vehicle is already present in the detection zone. The manufacturer claims that count accuracies greater than 95 per cent are possible, depending upon the type and volume of traffic and loop configuration.

The detector, in and of itself, only has the capability of providing cumulative vehicle counts (i.e., the total number of vehicle actuation since the last time the count was reset to zero). The detector itself does not let the user define a specific count interval over which to accumulate vehicle detection. To obtain counts *over a specified interval*, a device must be connected to the secondary outputs of the detector units which monitors the detection accumulation over a

specified period. However, by tying the collection interval to the phase outputs, it may be possible to obtain performance measures such as the number of vehicles arriving during the red interval and the number of vehicles serviced during the green interval.

CHAPTER 4:
POTENTIAL PERFORMANCE MEASURES

INTRODUCTION

One objective of this project was to identify and define potential performance measures that could be used to evaluate the effectiveness of traffic signal timings. FHWA defines a good performance measure as one that has the following attributes:

- is accepted by and meaningful to customers;
- tells how well goals and objectives are being met;
- is simple, understandable, logical, and repeatable;
- shows a trend;
- is unambiguously defined;
- allows for economical data collection;
- is timely, and
- is sensitive.

Using a systems engineering approach, we developed a list of potential performance measures that could be used to evaluate the effectiveness of traffic signal timings at an intersection. The report *Serving the American Public: Best Practices in Performance Measurement* (*1*) recommends including the following in developing a performance measure:

- the specific goal or objective from which it is derived;
- the data required, frequency of measurement, and the data source;
- the calculation methodology, including required equations, and precise definition of key terms;
- reports in which the data will appear and the graphical representation that will eventually be used to display the data; and
- any other relevant rationale for the measure.

MEASURES OF RELIABILITY

Blanchard and Fabrycky (5) define reliability of a system as follows:

"… the probability that a system or product will accomplish its designated mission in a satisfactory manner, or in more specific terms, the probability that it will perform in a satisfactory manner for a given period when used under specified operating conditions."

Among other things, the "mission" or purpose of a traffic signal is to provide for the orderly assignment of right-of-way for conflicting vehicles at an intersection with minimal disruption to the continuous movement of vehicles. The signal timing at an intersection, particularly an isolated intersection, should be designed so as to maximize the number of vehicles that can get through the intersection while, at the same time, avoid stopping vehicles unnecessarily. Under the most ideal of situations, the best performing signal timing would be one whereby no vehicle would have to stop to wait for a signal indication. In the real world, however, it is simply not possible to design a signal timing plan where no movement would ever have to stop. The general philosophy used to set the signal timing at most intersections is to keep the signal green for the through movements of the major roadway as much as possible (in the hopes of maximizing the total number of vehicles through the intersection) and make all other movements (such as the cross-street movements and turning movements) stop at the intersection. The controller terminates the major street green indication only when there is either 1) a large enough gap between successive vehicles on the major street to allow vehicle to stop safely; or 2) the minor cross-street movements are having to wait too long to be serviced. To measure the effectiveness of the signal timing plan to achieve these objectives, the following performance measures are proposed:

- the average number of times a phase was activated in a given evaluation period,
- the average number of vehicles served per cycle during a given evaluation period,
- the average number of vehicle stored per cycle during a given evaluation period,
- the probability of a vehicle having to stop at an approach during the evaluation period, and
- the percentage of overloaded cycles (or cycle failures) during a given evaluation period.

Average Number Phase Activations

One of the first things to look at when assessing the effectiveness of the signal timing at an intersection would be the frequency at which a phase was activated during an evaluation period. The average number of times a phase was activated during the evaluation period is a measure that allows the traffic engineer or technician to determine the relative magnitude at which a problem or a situation is occurring. This measure essentially provides context to the other performance measures and is needed to compute many of the other performance measures listed below.

To obtain this performance measure, a counter would need to be set up in a system that would increment each time a phase became active. One counter would be required for each phase that was serviced at the intersection. The system would monitor all the "PHASE ON" outputs from the controller and increase the counter each time the PHASE ON output changed its state from "OFF" to "ON".

Average Number of Vehicles Served per Cycle

The number of vehicles served is equivalent to the volume being served by the signal during each phase on an approach. Data for this measure can be collected by counting the number of vehicle actuations received by a detector when the particular phase serving that approach is active. The number of vehicles served could be measured relatively easily by placing a small, directional detector downstream of the stop line. The detector would need to be set far enough downstream from the stop line so that overhanging vehicles would not trigger the detector, but not too far so that crossing or turning traffic would trigger an actuation. The average number of vehicles served per cycle can be calculated by dividing the total number of vehicles served by a phase during the evaluation interval by the total number of times that the particular phase was activated during the same interval.

To collect the data needed to compute this performance measure, the system should monitor the status of the PHASE ON output of the controller as well as the status of the stop line detectors. When the status of the PHASE ON output changes from "OFF" to "ON", the system should begin counting the number of vehicles served during that phase by initializing the "# of Vehicles Served" counter. A vehicle would be assumed to be served by the signal each time the status of the stop line vehicle detector(s) associated with the phase change their status

from "OFF" to "ON". The "# of Vehicles Served" counter should be incremented each time a vehicle is detected. When the phase is over (i.e., the Phase On output changes its status from "ON" to "OFF"), the system should store the value of the "# of Vehicles Served" for retrieval at the completion of the cycle.

Average Number of Vehicles Stopped per Cycle

The exact number of vehicles stopping per cycle is a difficult parameter to measure because today's technology does not permit the tracking of individual vehicles. However, the number of vehicles stopping on an approach can be estimated by looking at the number of vehicles that arrive at the intersection during each red interval using a detector located upstream of the intersection. As long as the detector was relatively close to the intersection, vehicles detected arriving on the approach during the red interval could be assumed to have stopped at the intersection. To ensure that the counter does not miss vehicles, the detector would have to be located far enough upstream of the intersection to prevent queues from backing up over the detector.

To collect the data required to compute this performance measure, a counter would need to be set up to count vehicles that arrive over the detector. This counter would be active only when the signal for that approach is displaying a red indication. Each time the state of the detector changes from "OFF" to "ON", the counter would need to be incremented. Once the green indication was displayed on the approach, the counter would cease counting vehicles and the value recorded. The counter would need to be reinitialized to zero at the beginning of each cycle (i.e., when the PHASE ON output from the controller changed its state from "OFF" to "ON").

Proportion of Vehicles Having to Stop on an Approach

One of the major objectives of generating good signal timing plans are to provide for the continuous movement of traffic and to minimize the number of vehicles that have to stop at an intersection. Therefore, one way to assess the effectiveness of a signal timing plan is to monitor the proportion of vehicles that are arriving on an approach to an intersection that have to stop. If the proportion of the vehicles arriving at an intersection that are required to stop is large (especially on a main-street major movement approach), then the traffic engineer/technician might look for problems with detection systems, the gap times, or the phase splits. On the other

20

hand, if only a small proportion of the arriving vehicles have to stop, then the engineer/technician might judge the signal to be operating effectively. This performance measure can be computed by taking the ratio of the number of vehicles arriving on an approach during the red interval to the total number of vehicles served on the approach during that phase.

To compute this performance measure, the system would be required to count both the number of vehicles that had to stop on an approach during each cycle and the total number of vehicles that are served during the cycle on that approach. Both of these data elements have been discussed above.

Percentage of Overloaded Cycles (Or Cycle Failures)

The *Canadian Capacity Guide for Signalized Intersections (6)*, hereafter referred to as the *Canadian Capacity Guide*, contains a procedure for determining an "overload factor." This overload factor determines the "proportion of cycles in which the accumulated demand exceeds the capacity of a given lane [or approach]." It is determined by computing the ratio of the number of overloaded cycles to the total number of consecutively surveyed cycles. The equation used to compute the overload factor is as follows:

$$OF = \frac{N_o}{N} \times 100\%$$

Where,

$OF =$ the overload factor

$N_o =$ the number of overloaded cycles

$N =$ the total number of consecutively surveyed cycles, with $N \geq 20$.

The *Canadian Capacity Guide* defines an overloaded cycle to be one in which 100 per cent of the green time was utilized by traffic and that there was a queue present at the end of the amber period. In determining if the phase is overloaded, the procedures specified in the manual require that the phase be divided into sub-intervals (usually 5 seconds) and that an observer determine 1) if all of the phase sub-intervals are utilized by at least one vehicle, and 2) if there was a queue present at the end of the amber period.

A similar approach could be used to determine the percentage of overloaded cycles, with a slight variation. The *Canadian Capacity Guide* approach would work well for a phase of a fixed duration, but under actuated control, if the controller detects a gap in the traffic stream of a

certain size, the controller will terminate the phase. If the controller was functioning properly, then the intersection would never have any portion of the phase that went unutilized – instead the controller would gap the signal out and go to the next phase. The only time that a phase could be considered overloaded is when the following two conditions are met:

- when the phase terminates as a result of it either a)reaching its maximum limit or b) being forced off, and
- a constant call for service existing the entire time that the phase was active.

To determine if a phase terminated with demand, the system would need to monitor the status of the phase call detectors during the entire duration of the phase. If, at the start of the red clearance interval, the system detected a call on the phase detector AND the phase call detector never changed its state (from "ON" to "OFF") for the entire time the phase was active, then one can assume that the queue did not clear. When this occurs, the system need to flag the phase as experiencing a cycle failure. The system could then compute the performance measures by dividing the number of times a phase was flagged as experiencing a cycle failure by the total number of times that the phase was activated during the evaluation period.

MEASURES OF EFFICACY

Efficacy is defined as the "power or capacity to produce the desired effect; the ability to achieve results"(7). In signal timing, the desired effect is to respond to calls for service as quickly and as safely as possible without generating excess delays. For signal timing, efficacy can be measured using the following parameters:

- the average cycle time,
- the average duration of each phase,
- the average time required to service a call, and
- the average proportion of the green interval that was used to service the queue from the preceding red interval.

Each of these performance measures are discussed below.

Average Cycle Time

Another measure that might be useful to a traffic engineer or technician evaluating the effectiveness of the signal timing at an intersection would be average time duration between servicing each phase. For the purposes of this research, cycle time is defined as the time differential between the start of one phase in the current cycle to the start of the same phase in the previous cycle. Mathematically, cycle time can be expressed by the following equation:

$$CT = t_{\phi x,n} - t_{\phi x,n-1}$$

Where,

$\quad CT =$ Cycle Time (sec)

$\quad x_{\phi x, n} =$ Start of phase x (Φx) for the current cycle, n

$\quad x_{\phi x, n-1} =$ Start of the same phase x (Φx) for the previous cycle, n-1

For pre-timed signals, cycle time is equivalent to the cycle length. This is because with pre-timed signals, the start time of each phase occurs at the same point every cycle. But with fully actuated signals, cycle time is not the same thing as cycle length. With fully actuated control, the duration (and, to some degree, the sequencing) of each phase can vary from cycle to cycle. Cycle time measures these potential fluctuations and provides operators with an idea of the relative length of time between servicing each phase. Approaches that have moderate to light demand and/or sporadic arrival patterns would exhibit long cycle times. Approaches that experience very uniform or heavy demand would likely exhibit short cycle times. Long cycle time could also be an indication that the maximum (or MAX) timers in the controller may be set too long.

To collect the data needed to produce this performance measure, the system would need to measure and retain a timestamp for when each phase in the cycle became active. The system would monitor the PHASE ON outputs for each phase and then timestamp when this output changed its state from "OFF" to "ON". The average cycle time for each phase could then be computed by averaging the duration of each cycle time for a phase over the total number of times a phase was activated during an evaluation period.

Average Phase Duration

One measure that can be used to examine the effectiveness of a traffic signal timing plan is to examine the average duration of each phase. With modern actuated controllers, the duration of any phase can vary each cycle between its programmed minimum and maximum durations, depending upon the demand on the approach. A phase that does not exhibit much demand and is not heavily utilized would have a tendency toward shorter phase durations (i.e., close to either the sum of the minimum green interval, the yellow change interval, and the red-clearance interval or the minimum green interval plus the pedestrian clearance interval). Phases that have heavy demand or are highly utilized would tend to exhibit an average phase duration that is approaching the maximum phase duration coded into the controller. By examining the average phase duration at an approach, the operator would be able to determine how much of the allowable green time provided on an approach was being used each cycle. For example, if the average phase duration was relatively high, and yet there was not much demand on the approach or if time to service calls on the other approaches is relatively high, this might be an indication that the extension setting in the controller might be too long, allowing a phase to be extended by straggling demand.

The phase duration can be determined by measuring the time differential between when the "PHASE ON" output pin from the controller changed it state from the "ON" condition to the "OFF" condition. Mathematically, the formula for computing phase duration can be expressed as follows:

$$Phase\ Duration_{PhaseX} = t_{PHASEONX="OFF"} - t_{PHASEONX="ON"}$$

Where,

$Phase\ Duration_{Phase\ X} =$ duration of phase for Phase X

$t_{PHASEONX="OFF"} =$ the timestamp for when the state of the PHASE ON output for Phase "X" changed its state from "ON" to "OFF"

$t_{PHASEONX="ON"} =$ the timestamp for when the state of the PHASE ON output for Phase "X" changed its state from "OFF" to "ON"

Average Time to Service

One common complaint that signal operators tend to hear frequently deals with the amount of time that drivers experience waiting for the signal to turn green. Motorists generally equate wait time with delay. The longer they have to wait at an intersection, the higher their individual delay. Most motorists are willing to accept some wait time, as long as 1) the wait time is not excessive, and 2) they can see a reason for waiting. If they have to wait too long, some motorists might be tempted to disregard the signal indication and enter the intersection on a red indication.

Time to service is time differential between when a call for a phase came in to the controller and when that call was serviced by activating the phase. Time to service is equivalent to the maximum amount of time that a motorist would have to wait on an approach. It is a measure of the "snappiness" of the signal timing at an intersection. Intersections that are operating efficiently (or "snappy") tend to have lower times to service [i.e., less time between when a vehicle arrives at an intersection and when it is serviced by the signal (in the absence of demand on the opposing approaches)]. Signals that experience long times to service tend to have a tendency to increase driver frustration, particularly if there is little demand on the cross street.

Time to service is more applicable to isolated signals that are running in the actuated mode; however, time to service can also be used with fixed time or coordinated signals. For signals operating in these modes, time to service can be used as an indication that the cycle length is too long or that the phase splits on competing phase are too long.

Figure 1 shows the concept of time to service call. Time to service is determined by measuring the elapsed time from when the controller first receives a call for a phase to when the green indication is provided by the signal. It is the time differential between when the call for a phase first came into the controller to when the controller was about to service the phase (i.e., give it a green indication). To compute this measure, the following data are needed:

- a timestamp of when the controller first received a call for a phase on an approach, and
- a timestamp of when the controller began timing the phase (i.e., when the phase changed from a red indication to a green indication).

Note the requirement that the state of the phase call detector must remain in the "ON" state until the signal turns green for the corresponding approach. This requirement is needed to

filter out calls that are associated with vehicles making a right-turn-on-red and multiple loop detectors. The requirement assumes that those vehicles controls the signal will place a constant call to the controller.
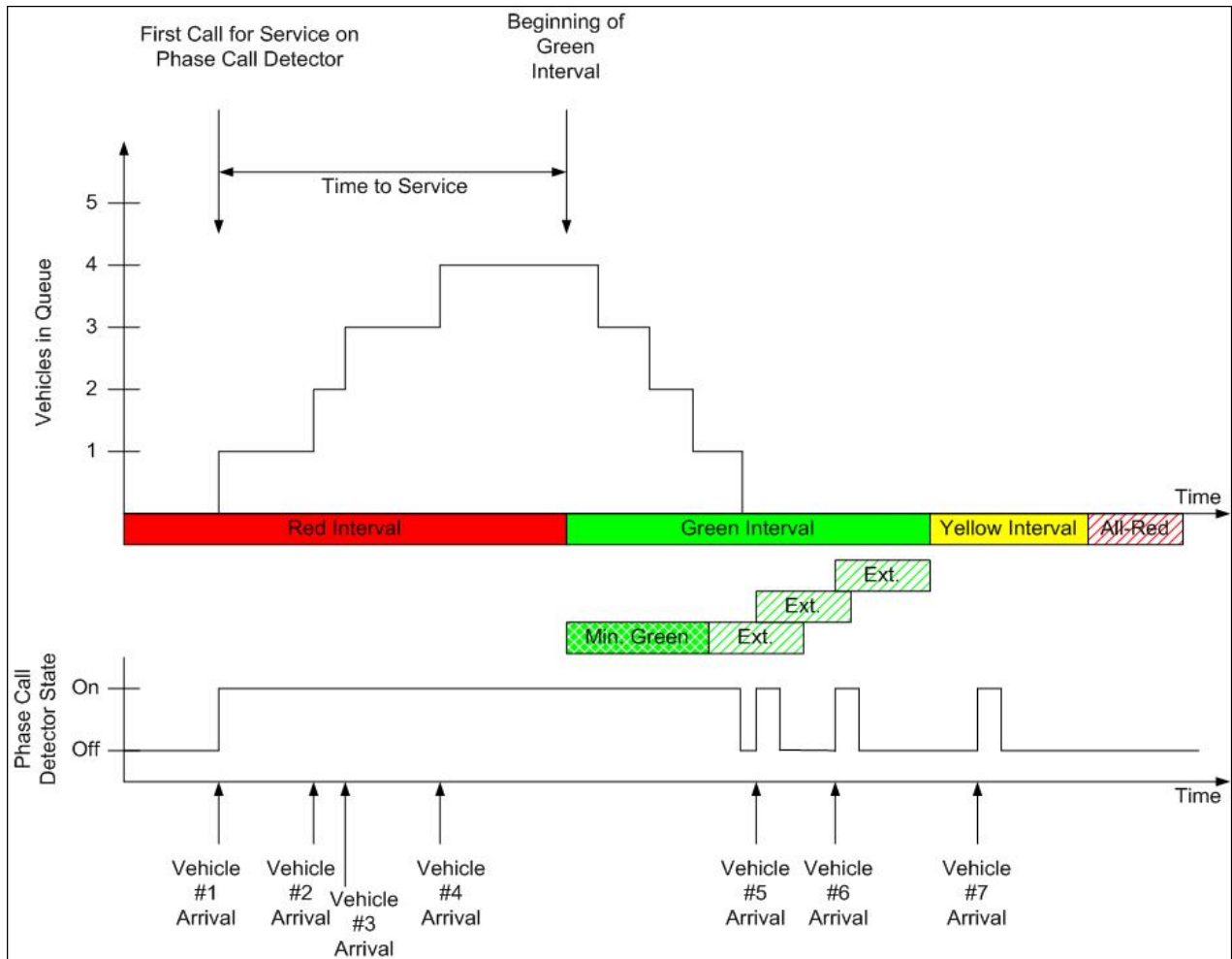

Figure 1.   Concept of "Time to Service" Performance Measure.

Average Proportion of Green Used to Service Queue

The *Highway Capacity Manual (HCM)* (8) defines the queue service time as the time needed by the signal to clear the demand (or queue) that has accumulated during the red interval leading up to the green indication. Engineers and technicians can use the proportion of the green interval used to service the queue to assess the effectiveness of the traffic signal timing.   To promote "snappy" operations, many traffic engineers and signal technicians will try to maximize the proportion of the phase that is used to service the queue by timing the traffic signal with

26

relatively short gap times.   An approach that experiences long gap times or long maximum times would tend to have a low proportion of the green used to service the queue, because random arrival of vehicles would have a tendency to continuously extend to the operations of the signal on that approach.

As shown in Figure 2, the queue service time is defined as the duration from when the signal turned green on approach to when the queue cleared at the intersection. The queue can be assumed to have cleared of the intersection once detection system ceased measuring a constant call on the associated phase call detector.
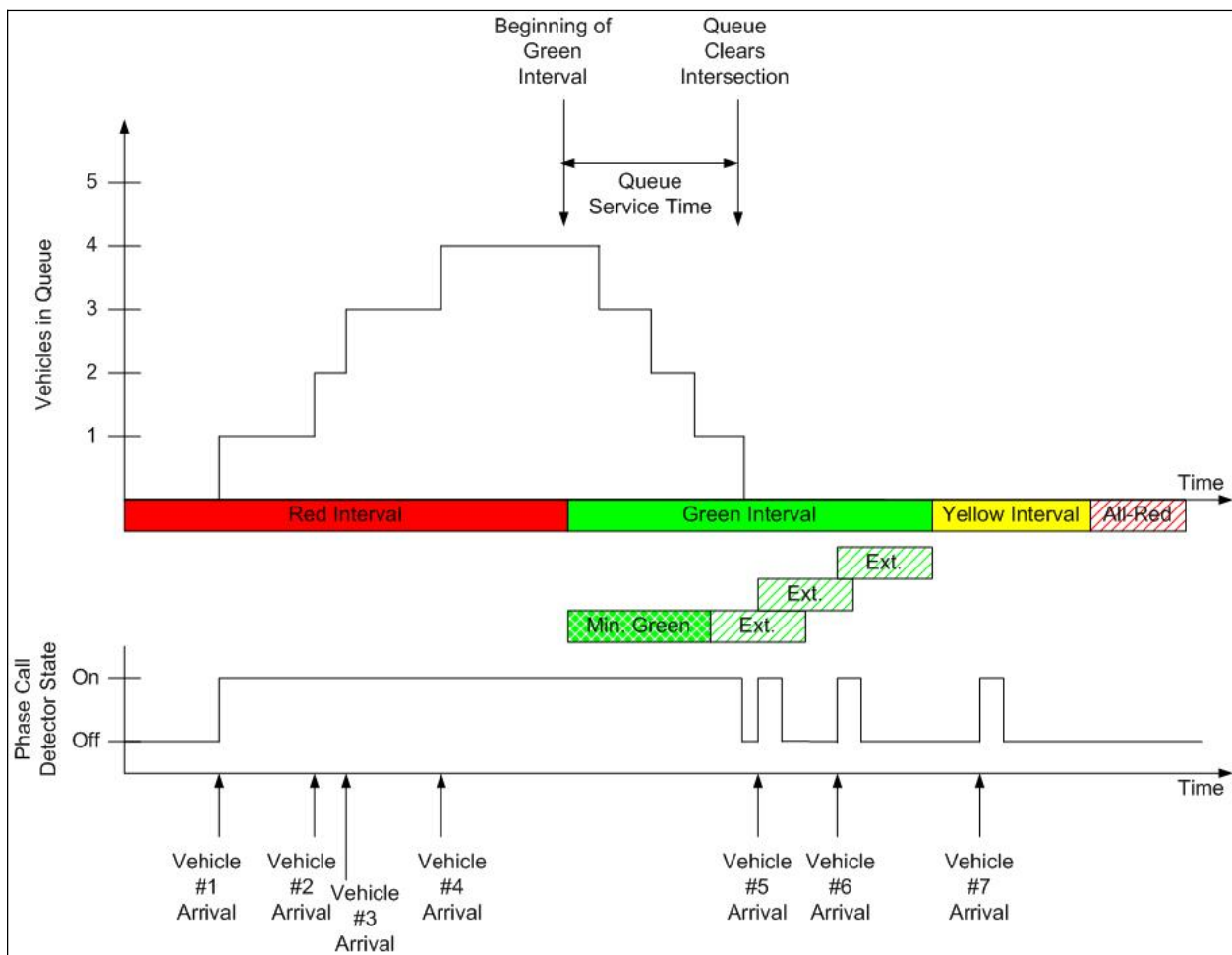

Figure 2. Concept of "Queue Service Time" Performance Measure.

The performance measure would be computed by taking the ratio of these two parameters.

27

For the purposes of this research, queue service time is defined as the time between when a phase becomes green and when the queued traffic clears the intersection. Measuring when the queue clears the intersection requires the use of a long loop detector operating in the presence mode located at the stop bar. If the loop is long enough, a queue over the detector is likely to place a constant call (or remain in the "ON" state) to the controller until the queue has cleared the detector. Therefore, any subsequent changes in the detector's state (i.e., from "ON" to "OFF") can be attributed to vehicles arriving at the intersection after the queue has cleared.

Using this assumption, queue service time can be estimated by measuring the time that elapses from when the green interval of a phase starts to when the queue clears the intersection. A queue is assumed to be present at the intersection as long as there is a call present on the phase detectors. Once the detector changes state (from "ON" to "OFF") the queue is assumed to have cleared the intersection. Queue service time can be expressed using the following equation:

$$QST = t_{Queue\ Clears} - t_{Phase\ X\ Green\ Starts}$$

Where

$\qquad QST =$ Queue Service Time (sec)

$\qquad t_{Queue\ Clears} =$ time queue clears intersection = time when phase detector first changes state from "ON" to "OFF"

$\qquad t_{Phase\ X\ Green\ Starts} =$ the time that the state of the Phase Green output pin for Phase "X" changed its state from "OFF" to "ON"

MEASURES OF SAFETY

A poorly timed traffic signal can potentially lead to an increased accident frequency and a high rate of motorists disregarding the signal indications. Generally, traffic engineers and technicians primarily use crash experience at an intersection as an indicator of poor performance; however, many engineers and researchers believe that vehicle conflicts can be used to identify potential operational problems before the crash history.

At traffic signals, the greatest potential for conflicts occurs during the change intervals. Potential performance measures that can be used to identify conflicts at signalized intersections include the following:

- the average number of vehicle entering the intersection on an approach during the yellow change interval,

- the average number of vehicles entering the intersection on an approach when the signal is in the all-red clearance interval, and
- the percentage of cycle during an evaluation period where one or more vehicles were observed entering the intersection during a specified period after the beginning of the red interval.

Each of these potential performance measures are discussed below.

Average Number of Vehicles Entering on Yellow Clearance per Cycle

Measuring the number of vehicles that enter the intersection during the yellow clearance interval can provide the signal timing engineer or technician with important information about the effectiveness of the signal timing. A large number of vehicles entering the intersection on yellow might be an indication that the intersection is operating too "snappy" and that the gap timer or extension timer might need to be increased. A large number of vehicles entering the intersection during the yellow clearance interval might also be an indication that the offset to the intersection might need to be adjusted to increase the size of the progression band through the intersection.

To compute this performance measure, the system could use the detector located downstream of the stop bar detector to count the number of vehicles that passed over the detector during the yellow change interval. The counter would be activated only when the controller was in the yellow clearance interval. Measuring the number of vehicles that enter the intersection during the yellow clearance interval involves two steps. The first step is to determine when the signal is in the yellow clearance interval. This can be done either by monitoring the Phase Yellow output pin for each particular phase, or by monitoring the Code Status Bit outputs. Table 3 shows the configuration of Code Status Bits that indication that the controller is timing the yellow clearance interval.

Table 3. Output of the Code Status Bits Showing that Signal is in the Yellow Change Interval.

Code #	Bit Status			State Name
	A	B	C	
4	Off	Off	**On**	Yellow

Once it is determined that the signal is in the yellow clearance interval, loop detectors set close to the stop line can be used to count the number of vehicles that enter the intersection. This counter would only be active during the yellow interval.

Average Number of Vehicles Entering on Red Clearance Interval per Cycle

An operator might also want to examine the average number of vehicles that enter the intersection during the red clearance interval. This performance measure could be used to provide insight into whether the yellow and all-red clearance intervals are adequate. It could also be used to help direct enforcement efforts at the intersection.

To determine the number of vehicles entering the intersection during the red interval requires that a counter be activated when the controller is 1) timing all-red clearance interval, or 2) a fixed time interval after the onset of the red indication. A detector located just downstream of the stop bar can be used to determine if a vehicle has entered the intersection. Any actuations that were recorded on the detector when the red indication is on would be assumed to be from vehicles entering the intersection during the red clearance interval. To compute the performance measure, the system would need to be programmed to collect the following data:

- the total number of vehicles in the evaluation period entering the specified portion of the red interval, and
- the total number of cycles experienced during the evaluation period.

Determining this parameter involves using the stop line detector(s) and the output of the Code Status Bits in the controller to count the number of vehicles that enter the intersection during the red clearance interval. The first step is to determine if the current active phase is in the red clearance interval. This can be done by monitoring the Code Status Bit outputs for each ring. The controller is timing the red clearance interval when the Code Status Bits is showing that displayed in Table 4.

Table 4. Output of the Code Status Bits showing that Signal is in the Red Clearance Interval.

Code #	Bit Status			State Name
	A	B	C	
5	**On**	Off	**On**	Red Clear

Upon determining that the intersection is in the red clearance interval, the performance monitoring system shall begin monitoring the intersection count detectors for the phase that is active to vehicle entering the intersection. A counter that counts the number of vehicles entering the intersection on red should be incremented each time the intersection count detectors change their status from "OFF" to "ON". When the red interval is over (i.e., the Code Status Bits will change their state to some other value than shown above), the system should output and store the value of the counter. This counter value should be associated with the phase that just completed timing.

A second option is to allow the user to set the duration into the red interval the system should monitor the approach for vehicle entering the intersection. Using this alternative would require that the system be set up slightly different. The system would need to contain a count down timer that would be initialized with the output of the Phase Red status changed from "OFF" to "ON". The initial value of the count down timer would need to be either 1) hard-coded in the system or 2) allowed to be set by the user during the initialization of the system. As the timer is counting down, the status of the stop line vehicle detectors would need to be monitored and the "# Vehicles Entering On Red" counter incremented each time the detector changed its status from "OFF" to "ON" When the countdown timer has expired (i.e., its value had reached zero), the value of the counter should be stored by the system.

Percentage of Cycle Experience a Red-Clearance Violation

One potential real-time measure that can be used to evaluate the signal performance would be to track the violation rate of vehicle entering during the red-clearance interval (or red-light violation [RLV]). The RLV would measure percentage of cycles in a given period where a vehicle was detected entering the intersection 1) during the all-red clearance interval, or 2) a

fixed time interval after the onset of the red indication. The measure could be computed using the following equation:

$$RLV = \frac{N_{RLV}}{N_C} \times 100\%$$

Where,

RLV $=$ red-light violation rate

N_{RLV} $=$ number of cycles were one or more vehicle was observed entering the intersection 1) during the all-red clearance interval or 2) up to x seconds after the onset of the red interval on an approach where x $=$ the number of seconds after the onset of red that red-light violations are counted;

N_C $=$ total number of cycles occurring during a user-defined interval.

A high RLV value would imply that a significant number of cycles had potential red-light violations, and could be used to assess problems associated with dilemma zones, insufficient clearance interval, or could be used to identify the need for selected enforcement. This measure could also be used to track the effectiveness of a red-light-running enforcement program.

To compute this measure, the data needed are as follows:

- the number of cycles where one or more vehicle was detected entering the intersection up to x seconds after the onset of the red interval; and

- the total number of cycles that occurred during the evaluation period.

To determine the number of cycles where a red-light violation occurred, a counter would need to be set up in the data collection system. The counter would be activated at the onset of the red clearance interval for a phase and could be set up to remain active for a fixed number of seconds defined by the user (i.e., 0-5 seconds). A vehicle would be determined to have entered the intersection if the system detected a relatively short duration between a detector "ON" and detector "OFF" pulse from a detection zone at or near the stop line of the approach.

The data source needed to compute this final performance measure is the number of cycles that vehicles were detected entering the intersection during the red interval. This involves monitoring the controller and a detector downstream of the stop line to determine if the detector changed its state (i.e., from "OFF" to "ON" and then again to "OFF" in a relatively rapid

32

succession) when the phase indication is in the red clearance or red interval. If this sequence of events occurs, the cycle should be flagged as one where vehicles entered the intersection on red. The flag would be reset to "ON" when the green indication for the next phase changed its state from "OFF" to "ON". A counter could then be used to count the number of cycles that were flagged as having a vehicle enter the intersection during the red interval.

REFERENCES

1. FHWA Travel Management Web Page on Performance Measures
    (http://www.ops.fhwa.dot.gov/Travel/Deployment_Task_Force/perf_meaures.htm).
    Accessed October 2003.

2. Shaw, T. *Performance Measures of Operational Effectiveness for Highway Segments and
    Systems*. NCHRP Synthesis 311. National Cooperative Highway Research Program,
    Transportation Research Board, Washington, D.C., 2003.

3. *Autoscope Solo® User's Guide*. Econolite. Anaheim, CA, 2003.

4. *Oracle /2 Loop Monitor™. Operation Manual*. Eberle Design, Inc. Phoenix, AZ.

5. Blanchard, B.S., and W.J. Fabrycky. *Systems Engineering and Analysis*. Third Edition.
    Prentice Hall, Upper Saddle River, NJ, 1998.

6. Teply, S. D. Allingham, D. Richardson, and B. Stephenson. *Canadian Capacity Guide for
    Signalized Intersection*. Second Edition. Committee on the Canadian Capacity Guide for
    Signalized Intersections, Institute of Transportation Engineers, District 7, 1995.

7. *The American Heritage Dictionary of the English Language*. William Morris, Editor.
    Houghton Mifflin Company, Boston, MA, 1979.

8. *Highway Capacity Manual*. Transportation Research Board, National Research Council,
    U.S. Department of Transportation, Washington, D.C., 2000.