

---

# CHAPTER 2

## Freeway- Management Systems



Freeway-management systems consist of strategies/system components and technologies combined to monitor, control, and manage freeway traffic more effectively. Strategies/system components and technologies in use include: ramp control (e.g., ramp metering, ramp closure); freeway mainline metering; freeway-to-freeway metering; reversible roadway control (e.g., lane control, variable speed control); priority control for high-occupancy vehicles (HOV) (e.g., priority access control, HOV facilities); transportation management during reconstruction; surveillance and detection (e.g., vehicle detectors, call boxes, CB monitoring, weather and environmental detection, overheight vehicle detection, automatic truck warning system, closed circuit television (CCTV)); driver information systems (e.g., changeable message signs (CMS), lane-use control signals, highway advisory radio, call boxes and commercial telephone, in-vehicle systems); communications (e.g., media types, data and voice, video).

Freeway management includes a Freeway Management Center (or multiple centers where responsibility for the freeway system is shared by more than one operating entity in a metropolitan area) and links to other ITS components in the metropolitan area. From these centers, personnel electronically monitor traffic conditions; activate response strategies; and initiate coordination with intraagency and interagency resources, including emergency response and incident-management providers.

Closed-circuit television and an array of sensors (e.g., inductive loops, magnetometers, microwave radar, ultrasonic, infrared, video image processing, automatic vehicle identification, and passive acoustic devices) may be used to electronically monitor freeway conditions in real-time. Other sources of information concerning real-time freeway conditions include communications received from police and maintenance personnel, as well as cellular telephone reports called in from drivers.

Traffic condition data are analyzed to identify the cause of a flow impediment and to formulate an appropriate response in real-time. Traffic control devices, such as ramp meters or lane control devices, may be proactively applied to provide a better balance between freeway travel demand and capacity during congested conditions. Information may be provided to travelers through roadside traveler information devices such as dynamic message signs and highway advisory radio. Emergency response and incident-management providers may be notified to respond to nonrecurring incident events. Automatic Vehicle Identification (AVI) readers may be used to acquire probe vehicle data.

#### SOURCES:

- *Traffic Control Systems Handbook*, February 1996, Publication No. FHWA-95-032, Prepared for FHWA by Dunn Engineering Associates.
- *Freeway Management Handbook*, available spring 1997 FHWA Publication.
- *Inform Evaluation, Vol. 1: Technical Report*, 1991, Publication No. FHWA-RD-91-075.
- *Inform Evaluation, Vol. 2: Executive Report*, 1991, Publication No. FHWA-RD-91-076.
- *Assessment of Advanced Technologies for Relieving Urban Traffic Congestion*, December 1991, NCHRP Report 340.

## 2.1 RAMP CONTROL

Entrance ramp control is the most widely used method of freeway traffic control. Its main objective is to limit the number of vehicles entering the freeway so that demand does not exceed capacity. The techniques/strategies for ramp control include ramp metering and ramp closure.

### 2.1.1 TECHNIQUES/STRATEGIES

#### Ramp Metering

Ramp metering is a method of regulating or restricting traffic flow (limits the rate at which traffic can enter the freeway) in such a manner that capacity on the freeway, downstream of the on-ramp, is not exceeded. The calculation of metering rates depends on the metering purpose: congestion reduction or merging operation safety.

- a. *Congestion reduction* - If the metering system intends to reduce congestion, demand must be kept less than capacity. Metering rate is based on the relationship between upstream demand, downstream capacity, and the volume of traffic desiring to enter the freeway at a particular ramp.
- b. *Safety* - In many cases, ramp metering provides a smoother ramp merging operation. The primary merging safety problem involves rear-end and lane-change collisions caused by platoons of vehicles on the ramp competing for gaps in the freeway traffic stream. Metering breaks up these platoons and facilitates single-vehicle entry.

The primary concerns of ramp metering are:

- Development of queues that back up onto the surface streets.
- Potential diversion of freeway trips to adjacent surface streets.
- Potential for increased violations of the ramp signal.
- To the public, ramp meters are often seen as a restraint on a roadway normally associated with a high degree of freedom—a new form of control where before there was none
- Even though a 2-to 5-minute wait at the ramp is more than compensated for by improved travel time on the freeway, motorists may only notice the waiting time on the ramp.
- A perceived advantage to longer distance commuters at the expense of shorter distance travelers. Close-in residents, for example, may be deprived of immediate access to the freeway, while long-distance commuters can enter beyond the metered zone and receive all the benefits of ramp metering without ramp delays.

#### Modes of Ramp Control

There are two basic Modes of Ramp Control: restrictive and nonrestrictive metering.

- a. *Restrictive Metering* - By in this mode, the metering rate may be used to measurably reduce the average demand in order to prevent main roadway congestion or to alleviate the extent of existing congestion. The two constraints placed upon restrictive metering are:
  - 1. The minimum metering rate that may be practically implemented while still retaining reasonable driver compliance (approximately 200 to 250 vehicles per hour); and
  - 2. Available ramp queue storage space.
- b. *Non-restrictive Metering* - By this mode, the metering rate is increased to clear that portion of the queue extending beyond available ramp storage.

In general, those ramps with adequate storage and one or more viable alternate routes will operate restrictively much more frequently than those with inadequate storage and poor alternate routes.

### Entrance Ramp Control

There are three categories of entrance ramp control: pretimed, local actuated or traffic responsive, and system control.

- a. *Pretimed* - The simplest form of control, whereby the metering rates are fixed and change only in accordance with a preset time-of-day/day-of-week schedule (characterized by time-clock rather than traffic-responsive operation).

This type of ramp metering is always recommended as a backup in the event of a communications, central computer, or other critical equipment failure.

- b. *Local Actuated or Traffic Responsive* - Ramp metering directly influenced by mainline traffic conditions. As occupancy levels on the mainline change, the metering rates on the ramp are changed accordingly.
- c. *System Control* - Ramp metering of a system of entrance ramps based on total freeway real-time traffic conditions. There is an interconnection of ramps that permits the conditions at one location to influence the metering at the other locations.

### Ramp Metering Strategies

There are four basic strategies for ramp metering. These are clock time metering, demand/capacity metering, speed control metering, and gap acceptance merge-control metering.

- a. *Clock Time Metering* - A count down clock is assigned to each associated on-ramp and the signal is set to green each time the clock returns to zero.
- b. *Demand/Capacity Metering* - An evaluation of excess capacity immediately downstream of the metered on-ramp is performed at given intervals, based on counts from surveillance detectors on the freeway mainline.

- c. *Speed Control Metering* - The procedure for this form of ramp metering is similar to that of the demand/capacity strategy. A metering rate is established based on speed evaluations made at a freeway link detector station.
- d. *Gap Acceptance Merge-Control Metering* - Ramp vehicles are released by the control signal so as to merge smoothly in gaps, expressed in units of time and detected in the outside freeway lane.

#### SOURCES:

- *Ramp Metering Status in North America, 1995 Update, June 1995, Report No. DOT-T-95-17, Prepared by Gary Piotrowicz of FHWA and James Robinson of VDOT.*
- *Ramp Metering Design Guidelines, August 1995, Caltrans.*
- *Ramp Metering Policy Procedures, January 1995, Caltra.*
- *Route 85 Ramp Metering Plan, February 1995, Caltrans.*
- *Route 85 Ramp Meter Operation, Stage III, April 1995, Caltrans.*
- *Ramp Metering on Route 50 in Sacramento: First Year of Operation, Fall 1984, Caltrans.*
- *Evaluation of the Two-Capacity Phenomenon as a Basis for Ramp Metering, December 1990, Prepared by San Diego State University for Caltrans.*

#### Ramp Closure

The most restrictive form of ramp control is ramp closure. This involves closing the ramp to traffic on a permanent or a short-term basis. Ramp closure has been used successfully in numerous cities in the United States and Japan (e.g., Houston, Los Angeles, San Antonio, Fort Worth). This is appropriate where:

- Ramps have inadequate storage.
- Freeway traffic is operating at capacity on a section before the entrance ramp.
- Ramp does not allow traffic to merge into the freeway traffic stream without considerable hazard or disruption.
- Ramp introduces serious weaving problems.

## 2.1.2 TECHNOLOGIES

### Ramp Metering

#### Hardware

The following hardware technologies are used for the various categories of ramp metering.

a. *Left and Right Side Signals*

The most visible part of the ramp meter system is the ramp signal, usually pedestal mounted on both sides of the ramp. It consists of a three-section upper signal head and a two-section lower signal head.

b. *Signs*

There are two general types of signs used in ramp metering: warning signs and regulatory signs.

1. *Warning Signs* - The three main types of warning signs are:

- *“Signal Ahead” Warning Sign.* Recommended when the ramp meter signal or queue is not within the field of view of oncoming traffic. This type of sign is installed near the entrance to the metered ramp facing each lane of traffic entering the ramp.
- *“Meter On” Warning Sign.* This is an internally illuminated sign mounted on a vertical post. When illuminated, it alerts drivers that a ramp entry to the freeway is ahead. The “Meter On” sign may be used instead of the “signal ahead.”
- *“Prepare to Stop” Warning Sign.* Recommended on high-speed ramps or freeway connectors where the approach speeds exceed 45 mph.

2. *Regulatory Signs* - Several types of regulatory signs are available, including:

- *“Stop Here On Red” Sign.* This sign is located (mounted) below the signal indication and includes an arrow pointing down and to the right.
- *“One Vehicle Per Green” Sign.* This sign is located (mounted) approximately 50 ft. upstream of the stop bar.
- *“Merge Left” Sign.* This sign is usually installed at the point where a two-lane ramp begins to taper to a single lane. This sign is required whether a meter is present or not.
- *“Right Lane Buses and Car Pools Only” Sign.* These signs are installed at regular intervals in advance of the ramp meter if HOV bypass lanes are used on two-lane ramps. This is to indicate that the right lane is for HOV only.

c. *Demand, Output, Merge, and/or Queue Detectors on the Ramp*

1. *Demand Detector* - This type of detector is located just upstream of the stop bar on the entrance ramp. The signal will remain red until a vehicle is detected by the demand detector.
2. *Output Detector* - This detector is located just downstream of the stop bar on the entrance ramp. This detector is used to assure single vehicle passage.
3. *Merge Detector* - This detector is located in the merge area of the of the ramp and the freeway. This detector is used to detect vehicles attempting to merge into the freeway (detects presence of vehicles in the merge area).

4. *Queue Detector* - This detector is located on the ramp, well in advance of the ramp metering signal, near the frontage road. It is used to detect vehicles' presence for extended periods of time (an indication of long queues on the ramp). It is used at locations where ramp backups may impact surface street operation.

d. *Upstream, Downstream, and Bottleneck Freeway Mainline Detectors*

In traffic-responsive metering, upstream demand and downstream capacity, as well as real-time measurements at the bottleneck, are necessary to determine metering rates. Traffic variables including volumes, occupancies and speeds are measured using upstream, downstream and bottleneck mainline detectors and are used to determine the relationship between the demand and capacity at different positions before and after the on-ramp.

e. *Local and Master Controllers With Interconnect and Cabinet With Amplifiers*

The controller cabinet is the roadside enclosure that houses the ramp meter equipment including the controller (e.g., Model 170 or 179 controller). The most common types of cabinets used in ramp-metering operations are the California Type 334 cabinet and the New York Type 330 cabinet. The equipment housed in the cabinet includes:

1. *Model 170 or 179 Controller*. This controls traffic flow on the metered ramp.
2. *Input File, Model 222 Detector Amplifiers, Model 242 and Model 252 Isolators*. The input file is an electronic unit housing the input-detector ("loop") amplifiers and AC and DC isolators.
3. *Power Distribution Assembly (PDA)/Output File, Model 200 Load Switches*. This is an electronic unit housing the load switches.
4. *Auxiliary Output Device*. The auxiliary output file is an electronic unit housing optional Model 200 Load Switches for the extra signal heads required on a three-lane metered ramp.
5. *Terminal Boards*. These provide the interconnection for the controller modem to the Traffic Operations Center.

f. *Central Processor at Control Center With Data Communication Interface*

The equipment used in the Traffic Operations Center to monitor and control ramp meters includes:

1. *Freeway Operations Status Display (FOSD)*. This usually consists of a large screen display system that is visible to all operators of the freeway system. It is usually color-coded where the different colors are used to indicate traffic speeds in real-time.
2. *Central Processor*. The control processor monitors overall freeway conditions and is used to execute the ramp metering program.
3. *Communications Computer*. This controls the modem links between the Traffic Operations Center and the metered ramps.

4. *Computer Workstations.* Computer workstations are used to control the ramp metering central processor. They are used to input data or modify parameters of the ramp-metering program.

## Software

The following software technologies are used for the various categories of ramp metering:

- a. *Automatic Rate Selection Method.* Most traffic-responsive metering systems determine metering rate with this method. Metering rates are selected based on predetermined thresholds and rate tables.
- b. *Control Emulation Method.* This model, developed in Minnesota, emulates real-time selection metering using simulated freeway traffic.
- c. *Predictive Algorithm Method.* This method, based on statistical pattern recognition, anticipates bottlenecks one or two minutes prior to their occurrence. Time-series intervention analysis is used for implementation of the predictive algorithm method.
- d. *Off-Line Simulation Modeling.* Traffic simulation modeling is sometimes used by Traffic Operations Centers, off-line, to determine optimal metering rates based on simulations of different metering algorithms and their impacts on mainline and surface street operations. Such models include FREQ, INTRAS, FRESIM, CORFLO, FRED, etc. FRED (Freeway Ramp Evaluation Database) is used in the INFORM system.

## Ramp Closure

Available hardware technologies used for ramp closure include:

- Manually Placed Barriers (e.g., cross bucks, barrels, cones).
- Automated Barriers (e.g., gates similar to those used at rail-highway grade crossings).
- Warning Devices.
- Dynamic Message Signs.
- Signals.

## 2.1.3 CASE STUDIES

A listing of Freeway Management Systems that include ramp-metering operations installed around the country can be found in Appendix A <sup>1,2,3</sup>

### Portland, Oregon

The first ramp meters in the Pacific Northwest were installed along a 10-kilometer section of I-5 in Portland in January 1981. The meters are operated by the Oregon Department of Transportation.

I-5 is the major north/south link, and is an important commuter route through the metropolitan area. This initial system consisted of 16 metered ramps between downtown Portland and the Washington State line. Nine of the meters operated in the northbound direction during the evening peak and seven controlled southbound entrances during the morning peak. The meters operate in a fixed time mode. There are currently 58 ramp meters, operating on five different freeways.

Prior to metering, it was common along this section of I-5 for platoons of vehicles to merge onto the freeway and aggravate the already congested traffic. The northbound evening peak hour average speed was 26 kph. Fourteen months after installation, the average speed for the same time period was 66 kph. Travel time was reduced from 23 minutes (highly variable) to about 9 minutes. Premetered conditions in the southbound morning peak were much less severe hence the improvements were smaller. Average speeds increased from 64 to 69 kph resulting in only slight reductions in southbound travel times.

Other results evaluated for the evening peak period include fuel savings and a before-and-after accident study. It was estimated that fuel consumption, including the additional consumption caused by ramp delay, was reduced by 2,040 liters of gasoline per weekday. There was also a reduction in rear-end and sideswipe accidents. Overall, there was a 43 percent reduction in peak period traffic accidents.

#### Minneapolis/St. Paul, Minnesota

The Twin Cities Metropolitan Area Freeway-Management System is composed of several systems and subsystems that have been implemented over a 25-year period by the Minnesota Department of Transportation (MNDOT). The first two fixed time meters were installed in 1970 on southbound I-35 east, north of downtown St. Paul. In November 1971, these were upgraded to operate on a local traffic-responsive basis and 4 additional meters were activated. This 8-kilometer section of I-35 east has been evaluated periodically since the meters were installed. The most recent study shows that after 14 years of operation, average peak hour speeds increase from 60 to 69 kph, or 16 percent faster than before metering. At the same time, peak period volumes increased 25 percent due to increased demand, the average number of peak period accidents decreased 24 percent, and the peak period accident rate decreased 38 percent.

In 1974, a freeway-management project was activated on a 27-kilometer section of I-35 west from downtown Minneapolis to the southern suburbs. In addition to 39 ramp meters, the system included 16 closed-circuit television cameras, 5 variable message signs, a 2-kilometer zone of highway advisory radio (HAR), 380 vehicle detectors, and a computer control monitor located at the MNDOT Traffic-Management Center in Minneapolis. This project also included extensive "freeway flyer" (express bus) service, and eleven ramp meter bypass ramps for HOVs. An evaluation of this project after 10 years of operation shows that average peak period freeway speeds increased from 55 to 74 kph or 35 percent. During the same 10-year span, peak period volumes increased 32 percent, the average number of peak period accidents declined 27 percent, and the peak period accident rate declined 38 percent. More than one million dollars a year in road-user benefits are attributed to reduced accidents and congestion. This system also has positive environmental impacts. Peak period air pollutant emissions, which include carbon monoxide, hydrocarbons, and nitrogen oxides, were reduced by just under 2 million kilograms per year.

More than 300 additional ramp meters were implemented from 1988 to 1995, and there are currently 368 meters in operation. Further projects are now in the design and construction phases. During the next five years, the plans are to complete the ramp metering system that will cover the entire Twin Cities freeway network. The success of the Twin Cities system has shown that the staged implementation of a comprehensive freeway-management system on a segment-by-segment, freeway-by-freeway basis, over a long period of time, is an effective way of implementing an area-wide program.

### Seattle, Washington

In September 1981, the Washington State Department of Transportation (WSDOT) implemented metering on I-5 north of the Seattle Central Business District. Initially the system, named FLOW (not an acronym), included 17 southbound ramps that were metered during the morning peak, and 5 northbound ramps that were metered during the evening peak. Currently, the ramp metering system includes 54 meters on I-5, I-90, and SR 520. These meters are all operated by centralized computer control. Future expansion plans include additional ramp meters on SR-520 east of Lake Washington, all of I-405, and I-5 south of Seattle.

One evaluation of the initial 22-meter system showed that between 1981 and 1987, mainline volumes during the peak traffic periods increased 86 percent northbound and 62 percent southbound. Before the installation of metering, the travel time on a specific 11-kilometer course was measured at 22 minutes. In 1987, the travel time for the same course was measured at 11.5 minutes. During the same six-year period, the accident rate decreased by 39 percent.

A somewhat unique application of metering was implemented in Seattle on SR-520 in 1986. While diversion caused by metering is often controversial, one of the objectives of metering SR-520 was to reduce commuter diversion through a residential neighborhood. The meters were installed on the two eastbound ramps on SR-520 between I-5 and Lake Washington. One of these ramps, the Lake Washington Boulevard on-ramp, is the last entry onto SR-520 before the Evergreen Point Floating Bridge. Because there were no bottlenecks downstream of this ramp, traffic would normally flow freely onto the bridge and beyond. Motorists, especially commuters from downtown Seattle, were using residential streets to reach the Lake Washington Boulevard on-ramp to avoid congestion on SR-620. This on-ramp, however, was a major contributor to congestion on SR-520 because of the high entering volumes. By metering the ramp, it was anticipated that traffic diverting through the adjacent neighborhood from downtown would be discouraged by the delay caused by the meter. Motorists would instead use the Montlake Boulevard on-ramp which was also metered at the same time. An HOV bypass lane was also installed at the Montlake Boulevard on-ramp. Two other objectives of this project were to improve flow on SR-520 and to encourage increased transit use and carpooling.

After four months of operation, an evaluation of this two-ramp meter "system" indicated a 6.5-percent increase in mainline peak period volume, a 43-percent decrease in the volume on the Lake Washington Boulevard on-ramp, an 18-percent increase the volume on the Montlake Boulevard on-ramp, and a 44-percent increase in HOVs using the Montlake Boulevard on-ramp. Another indication of the effectiveness of the combination of the HOV bypass and the improved SR-520 flow is a decrease of 3 minutes in METRO (King County Department of Metropolitan Services) transit travel times for buses traveling from downtown to the east, and a 4-minute decrease for buses traveling

from University District to the east. The reliability of the bus travel times also improved and METRO adjusted the schedules for these routes accordingly.

In 1993, the WSDOT implemented weekend ramp metering for the first time. Three ramps north of Seattle on southbound I-5 have been metered several hours due to heavy weekend volumes. Because of this success, in March of 1995, weekend metering was expanded to include four additional southbound ramps.

In April of 1995, WSDOT began operating seven southbound I-5 meters during the evening commute. This is WSDOT's first metering implementation in both directions of a corridor during the same peak period. The motivation behind this operational change is that the traditional reverse commute direction has become increasingly congested. Prior to this, metering along this section had operated southbound (inbound toward Seattle) during the morning commute and northbound (outbound) during the evening commute.

#### Denver, Colorado

The Colorado Department of Transportation activated a pilot project to demonstrate the effectiveness of ramp metering on a section of northbound I-25 in March 1981. The initial system consisted of five local traffic-responsive metered ramps operated during the morning peak on a 4.7 kilometer section of I-25 south of the city. Periodic evaluations revealed significant benefits. An 18-month study showed that average peak period driving speed increased 57-percent and average travel times decreased 37-percent. In addition, incidence of rear-end and sideswipe accidents declined 5 percent due to the elimination of stop-and-go conditions.

The success of the pilot project led to expansion of the system. In 1984 a central computer was installed and a system coordination plan that permits central monitoring and control of all meters. Since 1984, additional ramp meters have been added and today there are 28. In late 1988 and early 1989, a comprehensive evaluation of the original metered section was conducted. A number of changes occurred between 1981 and 1989, the most significant of which was the completion of a new freeway, C-470, which permitted more direct access to I-25 from the southwest area and generated higher demand for I-25. Volumes during the two-hour morning peak period increased from 6,200 vph in 1981 to 7,350 vph in 1989 (on three lanes). Speeds measured in late 1988 decreased from the original evaluation, but remained higher than the speeds before metering was implemented; 69 kph before, 85 kph after in 1981, and 80 kph in late 1988. The frequency of accidents during the morning peak period did not increase between the original evaluation and 1989, as a result, the accident rate decreased significantly because of the increased volumes. Rear-end and sideswipe accidents decreased by 50 percent during metered periods.

An interesting unplanned "evaluation" of the system occurred in the spring of 1987. To accommodate daylight savings time, all of the individual ramp controllers were adjusted one hour ahead. Unfortunately, the central computer clock was overlooked. The central computer overrode the local controllers and metering began an hour late. Traffic was the worst it had been in years. This oversight did have a bright side for the Colorado Department of Transportation: since the incident, the media have been even more supportive of ramp metering.

In 1988, the Colorado Department of Transportation conducted a study to evaluate different levels of ramp metering control. The study compared ramp meters operating in local traffic-responsive mode versus meters operating under centralized computer control. The results showed that if local traffic-responsive metering could maintain freeway speeds of more than 90 kph, centralized control had little or no additional benefit. However, if local traffic-responsive metering was unable to maintain speeds near the posted speed limit of 90 kph, centralized control was very effective. Data showed speeds increased 35.5-percent, from 50 to 68 kph, and vehicle hours of travel were reduced by 13.1-percent. This evaluation shows the importance of implementing operating strategies that correspond to the needs of the freeway network.

#### Detroit, Michigan

Ramp metering is an important aspect of the Michigan Department of Transportation's (MDOTs) Surveillance Control and Driver Information (SCANDI) System in Detroit. The SCANDI metering operation began in November 1982 with six ramps on the eastbound Ford Freeway (I-94). Nineteen more ramps were added on I-94 in January 1984 and three more in November 1985. An evaluation performed by Michigan State University for MDOT determined that ramp metering increased speeds on I-94 by about 8-percent. At the same time, the typical peak hour volume on the three eastbound lanes increased to 6,400 vehicles per hour from an average of 5,600 vph before metering. In addition, the total number of accidents was reduced nearly 50-percent and injury accidents were down 71-percent. The evaluation done by Michigan State also showed that significant additional benefits could be achieved by metering the three freeway-to-freeway connectors on this section of I-94.

#### Austin, Texas

In the late 1970s, in Austin, the Texas Department of Transportation implemented traffic-responsive meters at three ramps along a 4.2-kilometer segment of northbound I-35 for operation during the morning peak period. This section of freeway had two bottleneck locations that were reducing the quality of travel. One was a reduction from three to two lanes and the other was a high-volume entrance ramp just downstream of a lane drop. Metering resulted in an increased vehicle throughput of 7.9-percent and an increase in average peak period mainline speeds of 60-percent through the section. The meters were removed after the reconstruction of I-35 eliminated the lane drop in this section. This situation shows the versatility of ramp metering in that it can also be used effectively as a temporary solution. Austin is in the preliminary design stages and is expected to begin ramp metering again in about three years.

#### Long Island, New York

At the other end of the spectrum from Austin's implementation is the INFORM (Information For Motorists) project on Long Island. The INFORM project covers a 34-kilometer-long by 8-kilometer-wide corridor, at the center of which is the Long Island Expressway (LIE). Also included in the system is an east-west parkway, an east-west arterial and several crossing arterials and parkways, a total of 207 kilometers of roadways. System elements include 70 metered ramps on the LIE and the Northern State/Grand Central Parkway.

In 1989 after two months of operation, an analysis of the initial metered segment was conducted. In the peak period, the study showed a 20-percent decrease in mainline travel time (from 26 to 21 minutes) and a 16-percent increase in average speed; from 47 to 56 kph. Motorists entering at metered ramps also experienced an overall travel time reduction of 13.1-percent and an increase in average speed from 37 to 45 kph. The maintenance of efforts (MOE) tests for this project include vehicle emissions. For this initial segment, the analysis indicates there was a 6.7-percent reduction in fuel consumption, 17.4-percent reduction in carbon monoxide emissions, 13.1-percent reduction in hydrocarbons, and 2.4-percent increase in nitrous oxide emissions. The latter is associated with the higher speeds. Initial observations of the effect of metering the INFORM project's four-lane parkway indicates the benefits may be even greater than those achieved on wider freeways. Intuitively this makes sense because the impact of an unrestricted merge on only two lanes (in one direction) can be severe.

A more extensive evaluation of the INFORM project was completed in 1991. Data from this study showed much more conservative results. It is believed that this study is more representative of the true traffic conditions. The main reason for this is related to the "queuing off" (shutdown of the meter due to excessive queuing) of the ramp meters. The original study did not include areas where metering was usually shut off due to heavy ramp volumes, while this study accounted for all ramps. This evaluation showed that while throughput had only increased about 2-percent, the average mainline speeds had increased from 64 to 71 kph, or about 9-percent. However, at two separate bottleneck locations, data showed increases of 53 to 84 and 53 to 89 kph, or gains of about 36 and 40-percent, respectively. This evaluation also included calculation of a "congestion index." This index is the proportion of detector zones for which speeds were less than 48 kph (30 mph). While no benefit was shown in the evening peak period, the morning peak period showed an improvement of 25-percent in the congestion index. The accident frequency also showed encouraging improvement with a 15-percent reduction compared to the control section.

San Diego, California

In San Diego, ramp metering was initiated in 1968. That system, installed and operated by the California Department of Transportation, now includes 134 metered ramps on 110 plus kilometers of freeway. No detailed evaluations of metering have been conducted on the San Diego system since the early installations, but sustained volumes of 2,200 vph to 2,400 vph and occasionally even higher, are common on San Diego's metered freeways. A noteworthy aspect of the program is the metering of eight freeway-to-freeway connector ramps. Metering freeway-to-freeway connectors requires careful attention to storage space, advanced warning, and sight distance. If conditions allow, freeway connector metering can be just as safe and effective as other ramp meters.

## 2.1.4 SUMMARY OF ENTRANCE RAMP METERING BENEFITS

Metering entrance ramps significantly improves mainline traffic flow. The preceding case study evaluations, as well as others, show metering consistently increases travel speeds and improves travel time reliability, both of which are measures of reduced stop-and-go, erratic flow. It should be emphasized that these benefits occurred even though in most instances mainline volumes had significantly increased. Metering helps smooth out peak demands that would otherwise cause the mainline flow

to breakdown. A strong case can be made from the data reported that metering actually increases the throughput of a freeway. The data from Minneapolis, San Diego, Seattle, Detroit, and Denver shows mainline volumes well in excess of 2,100 vph per lane on metered sections, and sustained volumes in the range of 5-percent to 6-percent greater than on pre-metered conditions. Improved traffic flow, particularly the reduction of stop-and-go conditions, also reduces certain vehicle emissions. This has been shown in both the INFORM project and in the Twin Cities Freeway-Management System.

The other direct benefit, but one that has not been fully quantified, is the reduction in accidents attributed to metering. The case studies presented in this report consistently show a reduction in accident rates of 24- to 50-percent. However, the benefits derived from accident reduction goes well beyond the direct costs related to medical expenses and vehicle damage. To illustrate, assume an accident blocks one lane of three at the beginning of the peak period on a freeway with a two-hour peak demand of 6,000 vph. Studies show an accident blocking one of three lanes reduces capacity by 50-percent. A 20-minute blockage would cause 2,100 vehicle-hours of delay, a queue more than 3 kilometers long, and take two-and-one-half hours to return to normal assuming there were no secondary accidents or incidents. Clearly the safety aspects of metering are a major benefit.

Ramp meters, or flow signals as they are called in Houston, have been used to improve traffic flow for years. Cities like Seattle, Minneapolis, and Portland all have similar systems that have reported measurable improvements in traffic flow. In Portland, studies showed that flow signals increased average freeway speeds by 40 kph in just 14 months. In Seattle, freeway travel time went from 22 minutes to slightly more than 11 minutes after flow signals were installed. In Minneapolis, freeway travelers experienced a 38-percent reduction in the peak-period accident rate after flow signals were installed.

## 2.2 FREEWAY MAINLINE CONTROL

Freeway mainline control is the regulation, warning, and guidance of freeway traffic in order to: achieve more uniform and stable traffic flow as demand for the facility approaches capacity; reduce the potential for rear-end collisions if congestion develops; facilitate incident management and recovery from congestion; divert freeway traffic to alternate routes to utilize corridor capacity; and change the directional capacity of the freeway by use of reversible lanes.

### 2.2.1 TECHNIQUES/STRATEGIES

#### Mainline Metering

Mainline metering is primarily used at locations approaching major traffic bottlenecks such as approaches to bridges or tunnels, and for regulating the number of toll booths that are open at any given time. By regulating the number of toll booths that are open at any given time, the traffic demand entering the freeway system via the mainline is controlled according to available freeway demand and capacity downstream.

Another application of mainline metering is metering of the HOV lane. Certain locations in California meter the on-ramp HOV lanes at the same locations as the other lanes. In Seattle, WA, there is one metered HOV lane (northbound on-ramp to the I-5 express lanes) that operates in conjunction with the Metro bus tunnel on-ramp. Mainline metering has also been implemented on the east end of the San Francisco-Oakland Bay Bridge, where signals are placed over each of the 14 nonpriority westbound lanes at a point downstream of the toll booths, releasing vehicles to merge into five lanes on the bridge. Where congestion on the freeway cannot be prevented by entrance ramp control, mainline metering is installed on the mainline approach to the entrance ramp.

Jacobson and Landsman suggest that mainline meters be installed on freeways approaching bottleneck locations where analysis indicates that improved traffic operations will result.

They also suggest guidelines for mainline metering systems, including:

- Whenever possible, install meters at locations on roadways that are level or have a slight downgrade, so heavy vehicles can easily accelerate.
- Install meters where the sight distance is adequate for drivers approaching the mainline meter to see the queue in time to safely stop.
- Provide an HOV lane that allows HOVs to bypass the mainline meter traffic queue. The HOV lane should extend from the mainline meter upstream to the rear of the worst anticipated traffic queue.
- Installation of mainline metering in Washington State would require an extensive marketing and publicity campaign.

## SOURCES:

- *Case Studies of Freeway-to-Freeway Ramp and Mainline Metering in the U.S.*, and Suggested Policies for Washington State, January 1994, by Eldon L. Jacobson and Jackie Landsman, TRB paper.

## Case Studies

a. *San Francisco*

The mainline metering facility on westbound I-80 at the San Francisco-Oakland Bay bridge is a unique system in that the meters are located just downstream of a 22-bay toll plaza. Westbound traffic approaching the San Francisco-Oakland Bay bridge passes through the toll plaza and is then metered so that the 22 lanes of traffic can be narrowed to four as efficiently as possible. HOT lanes are also provided that allow HOBS to bypass the traffic queues. Wait time at the meters can reach 30 minutes. However, with the meters off, queuing occurs on the bridge rather than at the meters, and the delays exceed those that occur when the meters are on. Public opinion of this metering has been quite good, considering the long delays. The mainline metering is operated during peak traffic periods, which vary, depending on traffic volumes. The meters are also operated outside peak periods whenever the traffic volume reached a preselected level or when an incident on the bridge blocks the traffic lanes.

b. *Hampton Roads Tunnel*

This tunnel, located in southeastern Virginia, provides the only connection between Hampton and Norfolk. Before mainline metering began in August of 1983, delays of up to two hours were not uncommon. The congestion in the tunnel caused cars to overheat and Carbon Monoxide (CO) levels to increase. The metering operation was commenced whenever traffic speeds in the tunnel dropped below 25 kph. As a result of metering, CO levels were reduced requiring less ventilation, and fewer vehicles overheated keeping lanes open and moving. The metering was an operational success. However, due to motorist complaints and a lack of political support, these meters have been removed.

c. *Baltimore Harbor Tunnel*

The Baltimore Harbor Tunnel was a test site used in the mid-1970s to study restricted flow on traffic facilities. The project involved implementing mainline metering 1,200 feet upstream of the tunnel portal. Metering was activated when flows became congested; this usually occurred at speeds between 32 and 40 kph. Results of the analysis showed increased speeds in the tunnel and capacity increases of about 10 percent per lane. Again, similar to the Hampton Roads situation, mainline metering was discontinued as a result of political pressure.

## Summary of Mainline Metering

Often new and different traffic engineering techniques like mainline metering take time to catch on. This relates to taking the "let somebody else make the mistake first" position. However, to study the effects of mainline metering, agencies can look at unregulated conditions that exist. Two examples of

unregulated mainline metering include, toll booths and incidents. Mainline toll booths act similar to meters in that there is a reduction in capacity because every vehicle must come to a stop. As seen in Chicago, after exiting the toll booths the freeway returns to freeflow conditions. Lane blocking incidents are also artificial reductions in capacity. Again, after passing the scene of the incident, traffic usually returns to free-flow conditions. It should be noted that even though the end results are similar to mainline metering, unregulated metering creates larger and unnecessary queuing. This is because these situations are not being managed in an organized manner, as would be the case with metering.

Freeway bottlenecks can occur for a number of reasons. Commonly, they occur at geometrically constrained locations where tunnels and bridges exist. At these situations, mainline metering is a viable traffic management tool that has been proven to increase throughput and improve flow conditions. However, mainline metering is the most restrictive form of ramp metering. For this reason it suffers from a lack of political and public support, making mainline metering difficult to implement (Gary Piotrowicz and James Robinson, 1995).

#### Freeway-to-Freeway Metering

Freeway-to-freeway ramp metering consists of installing signals, on the side of the roadway or overhead, on the ramps found at freeway-to-freeway interchanges. When freeway-to-freeway connectors remain uncontrolled, large traffic volumes have access to the freeway. These volumes are much higher than those found at typical freeway on-ramps.

The majority of the installations are found in California and Minnesota. In California, it was found that commuters are willing to wait in a queue for the meter in exchange for improvements in their trip farther downstream. Due to the public's increased frustration with traffic congestion, freeway-to-freeway metering was shown to have a high level of acceptance for traffic management. According to the Caltrans Ramp Meter Design Guidelines, the installation of ramp meters on connector ramps (freeway-to-freeway connectors) shall be limited to those facilities that meet or exceed the following geometric design criteria:

- Standard lane and shoulder widths.
- Tail-light sight distance, measured from 1,070-mm eye height to 600-mm object height, is provided for a design speed of 80 km/h minimum.
- All lane-drop transitions should be accomplished with a taper of 50:1 minimum.

Jacobson and Landsman suggest that meters be installed meters on freeway-to-freeway ramps where system performance and efficiency will be improved

They also suggest the following guidelines for freeway-to-freeway ramp metering systems:

- Consider and implement freeway-to-freeway ramp meters at locations where recurring congestion is a problem or where route diversion should be encouraged.

- Route diversion should only be considered where suitable alternative routes exist, to avoid diverting drivers through residential neighborhoods. Freeway-to-freeway ramp meters should be installed to improve the mainline flow and on-ramp merge, or to help multiple ramps merge into one ramp. If the intent of the meters is route diversion, then trailblazers or appropriate signing should be considered to educate drivers on preferred alternative routes.
- Avoid metering vehicles twice within a short distance.
- Avoid metering single-lane, freeway-to-freeway ramps that feed traffic into an add lane.
- Do not install meters on any freeway-to-freeway ramp unless analysis ensures that the main line flow will be improved, so that people using the freeway-to-freeway ramp are rewarded for waiting in line at a meter.
- Install meters on freeway-to-freeway ramps where more than one ramp merges before feeding onto the mainline, and congestion on the ramps occurs regularly. If traffic queues that impede mainline traffic develop on the upstream mainline because of a freeway-to-freeway ramp meter, then the metering rate should be increased to minimize the queues on the upstream mainline, or additional storage capacity should be provided.
- Freeway-to-freeway ramp meters should be monitored and be controllable by the appropriate traffic management center.
- Whenever possible, install meters at locations on roadways that are level or have a slight downgrade. Also, install meters where the sight distance is adequate.

## Case Studies

### *a. Los Angeles*

In Los Angeles, only a few interchanges have freeway-to-freeway ramps that are metered. The first was at the interchange of I-5 and I-110. Additional freeway-to-freeway ramp meters have been set up with the completion of the new Century freeway. In 1992, a two-lane ramp meter was installed on the connector of southbound I-5 to help manage the heavy traffic flow onto southbound I-110. The meter is turned on during the morning peak period.

### *b. Minneapolis*

Minneapolis has the largest number of freeway-to-freeway ramp meters in the U.S. Its two earliest installations were activated in 1971 at the interchange of Trunk Highway 36 and I-35E. Since then, 25 more freeway-to-freeway ramp meters have been activated throughout the Minneapolis/St. Paul metropolitan area. In 1974, ramp control was initiated from eastbound I-94, a six-lane freeway, to southbound Trunk Highway, a four-lane freeway to reduce the heavy congestion that was occurring downstream of this location and to improve the flow of the corridor in general. Before implementation, the two-lane ramp had carried 1,080 vph during the PM peak hour. After the meters had been turned on and both lanes had been metered, the volumes on the ramp dropped to 690 vph during the PM peak hour. MnDOT estimates that over two-thirds of the reduction are due to route diversion.

Typical delays on the ramp are over 1 minute, but they can react up to 8 minutes with one-quarter mile queues.

Four ramps were metered on the four-lane I-35W in 1975 as part of the Urban Corridors Demonstration Program in Minneapolis. The volumes on these ramps were reduced by approximately 20 to 25 percent after metering. Metering at this location is considered to be partly responsible for the 38 percent increase in speeds on northbound I-35W.

*c. San Diego*

San Diego began using freeway-to-freeway ramp metering in 1971 when it first installed a meter at the interchange of SR 15 and SR 94. Since then, three more installations have been added, two of which feed SR 94 and the third I-8. The justification for these meters was to better manage the queuing and delay that was occurring at these interchanges and to better manage the freeway system in general. The most recent direct connector metering installation in San Diego was in 1985 and was at the interchange of SR 67 and I-8. This three-lane meter was installed to relieve congestion on I-8 just downstream of the interchange. Before metering, this section of I-8 was frequently congested with long queues that extended upstream from the interchange. Since the meters were turned on, the flow downstream on I-8 has averaged 2,500 vehicles per hour per lane and speeds have averaged 60 mph.

In 1978, a meter was installed at westbound SR 94 to southbound SR 94 to relieve some of the congestion and queuing through the SR 94 interchange. In addition, a meter was added to the cross street feeding the southbound SR 125 on-ramp, which merges with the westbound SR 94 direct connector just before their confluence with southbound SR 125. Both meters are operated by an automated, traffic-responsive system on the basis of mainline volumes, so the meters are typically on only during peak periods. The direct connector also features a peak-period inside HOV lane in addition to the two regular lanes. All three lanes, including the HOV lane, are metered. When the meter on the direct connector was first activated, the ramp carried approximately 1,900 vph during the peak hour, with an average wait of 1 minute and a maximum wait of about 3 minutes. Today the volume on the three-lane ramp is approximately 2,900 vph and a maximum wait during the peak period can exceed 10 minutes. In spite of the high ramp delays, there have been very few complaints, and responses to the metering have continued to be positive (Eldon L. Jacobson and Jackie Landsman, 1994).

Summary of Freeway-to-Freeway Metering (Connector Metering)

Connector metering offers an additional opportunity for agencies to manage the flow of traffic on a freeway system. It also improves equity amongst motorists along the corridor by metering vehicles that may have entered the freeway on a un-metered entrance ramp. San Diego operates most of its connector meters during the morning peak, metering vehicles that predominately come from outlying areas.

In many instances meters are installed to improve mainline operations and better manage queuing and delay that occurs at freeway-to-freeway interchanges. However, as shown in Los Angeles, connector metering has other uses. The success stories described above all occurred while in

many instances long queues and delays occurred on the connectors. However, motorists have become tolerant because they experience mainline freeways that operate at high levels of performance (Gary Piotrowicz and James Robinson, 1995).

#### SOURCES:

- Case Studies of Freeway-to-Freeway Ramp and Mainline Metering in the U.S., and Suggested policies for Washington State, by Eldon L. Jacobson and Jackie Landsman, TRB paper, January 1994.
- Ramp Metering Status in North America, 1995 Update, June 1995, by Gary Piotrowicz and James Robinson, US Department of Transportation, Publication No. DOT-T-95-17.

## 2.3 DYNAMIC ROADWAY CONTROL

Messages reflecting closed lanes, reduced speed limits, and required lane changes are displayed for selected lanes of freeways on overhead mounted dynamic signals and signs. Lane-use control and variable speed control require signage over lanes of the freeway coupled with devices to detect speeds and volumes resulting in congestion. The detectors identify changes in traffic patterns that signal the beginning of congestion or incidents, thus prompting appropriate messages and lane and speed selection by the management center.

### 2.3.1 TECHNIQUES/STRATEGIES

#### Lane Control

Reversible lane control is used to change the directional capacity of a freeway in order to accommodate peak directional traffic demands. These are used where the peak period traffic volumes show a significant directional imbalance (e.g., 70 percent to 30 percent).

#### Variable Speed Control

Variable speed control is used to reduce the speed of traffic on a freeway to a level that corresponds to a maximum volume. As the peak flow demand increases, speed control can help improve the stability and uniformity of flow and reduce the occurrence of rear-end collisions. Also, lower speed limits could be used during hazardous driving conditions such as fog, rain, or snow.

### 2.3.2 TECHNOLOGIES

The following technologies are used for lane control and variable speed control:

These are overhead signals using indicators to permit or prohibit the use of special lanes on a freeway. Lane-use control signals are most commonly used for reversible-lane control. Other applications include: keep traffic out of a freeway lane to facilitate the merging of traffic from a ramp or other freeways; on a freeway near its terminus, to indicate a lane that ends; to indicate a temporary closure of a freeway lane due to an incident. The technology includes: fiber optic or light-emitting diode (LED); sign case with display elements (a downward green arrow, amber X, flashing amber X, and red X); power switch; photo sensor control to control light output levels.

This technique has been used on two of Dallas' busiest corridors to temporarily reassign traffic flow in specific lanes, thus doubling the number of lanes in the peak direction. The five-lane roadways dedicate four lanes per peak direction, and the fifth lane for the opposing traffic. These lanes are equipped with overhead dynamic message signs (DMSs) placed every 1,000 feet, alerting the motorist of the shift in lane direction by displaying an amber X in the lane about to change and a red X once the shift is complete. These installations have increased capacity by 33 percent for the peak direction. Types of DMSs used for this application include: fiber optic, LED, bulb matrix, blankout.

The Manual on Uniform Traffic Control Devices (MUTCD) describes the use of the lane control signals as follows.

- Downward green arrow - Used to indicate an open lane is open and that a driver is permitted to drive in the lane over which the arrow is located.
- Amber X - Used to prepare the driver to vacate the lane (similar to the use of amber indications at intersection traffic signals).
- Flashing amber X - Indicates that a lane may be used for a left turn (applicable to arterial streets only).
- Red X - Used to indicate that the lane over which it is displayed is closed to that direction of traffic, and that a driver shall not drive in that lane.

In a study performed by the Texas Transportation Institute on lane-use control signals at various sites in Texas, the results support the use of both the green arrow and the red X for these purposes. Nearly all subjects participating in the studies correctly interpreted the green arrow, and more than 80 percent correctly interpreted the red X. The studies also showed that the red X is perceived as requiring a reaction by motorists (to exit the lane) within 0.10 mile after it is displayed. However, the amber X was found to elicit widely varying responses depending on the context in which it was used. When displayed in conjunction with green arrows, motorist interpretations of the amber X were more consistent with its intended meaning of an impending lane closure and the need to vacate the lane. When presented in conjunction with a red X, motorists appeared less likely to associate an amber X with an impending lane closure. They instead interpreted the amber X to indicate congestion or other hazard in the lane downstream for which they would need to slow down and be careful. They further recommend that the amber X be replaced by an amber diagonal arrow to provide a more consistent interpretation of the action desired from the motorist.

### Variable Speed Limit Signs

These signs are of the same type as lane-use control (e.g., blankout signs the type used as pedestrian walk/don't walk, dynamic message signs).

### Gates

Hydraulically or electrically controlled lifting gates (barriers) present a physical barrier in the road and are used to control access to lanes on the freeway during recurring and nonrecurring congestion.

### Dynamic Message Signs

### Movable Lane Barriers

These are physical barriers installed, temporarily, by a transfer vehicle to separate lanes of a freeway. More discussion of these systems is found in Section 2.4.

## Regulatory Signs with Beacons

### Case Study

TransGuide - From the TransGuide ITS Design Report ([transguidewwwserver.datasys.swri.edu:80/articles/designreport.html](http://transguidewwwserver.datasys.swri.edu:80/articles/designreport.html) )

### Lane Control Signal (LCS)

The LCSs are the most numerous mechanisms for communicating with drivers. They provide limited, simple communications in the form of one signal for each lane of the freeway. The signal displayed for each lane can be a RED X, a GREEN DOWN ARROW, a YELLOW X, a YELLOW DOWN ARROW, a YELLOW SLANT ARROW pointing right, or a YELLOW SLANT ARROW pointing left. The RED X indicates that the lane is closed. The GREEN DOWN ARROW indicates that the lane is open. A YELLOW X indicates that the lane will be closed at some point downstream. A YELLOW DOWN ARROW indicates that the lane is open, but that there may be a hazard on an adjacent shoulder. A YELLOW SLANT ARROW pointing left or right indicates that the lane will be closed downstream and that the traffic should move into the lanes indicated by the arrow to find clear lanes or to exit from the freeway.

If a lane is blocked, the scenario implemented should have two RED Xs in that lane upstream of the incident. Before the two RED Xs, one YELLOW SLANT ARROW should be displayed. The contents of the LCS display are controlled by the TransGuide operators, either directly or through the initiation of scenarios. The LCS will normally display a GREEN DOWN ARROW for all lanes that are not blocked. The other signals are displayed based on specific scenarios initiated by the operators. In general, blocked lanes will have a RED X displayed by at least two LCSs in the lane upstream of the blockage. The use of a RED X on two LCSs ensures that the motorist will see a RED X in the lane even if one LCS fails. Upstream of a RED X, at least one signal in a blocked lane will normally contain a YELLOW ARROW advising motorists to begin moving out of the blocked lane. Hazards on the shoulders will normally cause a YELLOW DOWN ARROW to be displayed upstream from the hazard. Each scenario is evaluated to determine the best configuration of VMS messages and LCS signals.

The LCSs are composed of halogen bulbs, fiber optic bundles, color filters and lenses. Table 19 shows some of the characteristics of the specified LCS. Two halogen bulbs provide the illumination for each color and color filters are used to produce the appropriate colored light. Fiber optic bundles carry the light to the face of the LCS, and lenses direct the light toward the drivers. The lighted dots are surrounded by a flat black faceplate to provide sufficient contrast so the segments can be clearly distinguished at 1/4 mile. Each of the indicators is illuminated by bundles coming from two bulbs, so the indicators can still be seen even if one of the bulbs is burned out. The LCSs are required to provide automatic dimming for night operation and to ensure that dimming does not turn off a bulb if the other bulb providing that color is burned out.

There are several mechanisms to ensure that the LCSs are capable of operating continuously, reliably, and affordably. An LCS can be polled by the system and instructed to report the status of each lamp. Most LCSs can be visually checked using the camera system. The LCSs are required to automatically report failed bulbs. The specifications require easy service access to LCSs, and TransGuide specifica-

tions place specific environmental compatibility requirements on signals. Finally, the specifications also specify long life requirements for the bulbs. These measures should help ensure that the LCSs operate continuously, reliably, and affordably for many years.

SOURCES:

- *Driver Interpretations of Existing and Potential Lane Control Signal Symbols for Freeway Traffic Management*, 1993, TexasTransportation Institute, Research Report 1298-1.

## 2.4 PRIORITY CONTROL FOR HOV

Priority control provides preferential treatment for buses, carpools and other High-Occupancy (priority) Vehicles (HOV) that use the freeway. It is designed to relieve traffic congestion on freeways by encouraging better utilization of vehicles. This is accomplished through the following techniques/strategies.

### 2.4.1 TECHNIQUES/STRATEGIES

#### Priority Access Control

Priority access control is used in conjunction with entrance ramp metering and gives priority to HOV traffic by providing bypass lanes on the ramps or exclusive ramps. The HOVs avoid any delay associated with ramp queues and enter the freeway with no disruption. This priority access can be in the form of bypass lanes or a preferential metering rate. The priority lanes are usually restricted to buses or HOVs. The Freeway Management Handbook gives some design specs for these lanes: bypass lanes should be 12-foot wide with full ramp shoulders where possible, and should extend 300 feet beyond the metering signal to permit HOVs to merge with normal ramp traffic. The ramp bypass traffic should merge first with regular ramp traffic and then with freeway traffic. Regulatory signs with "Right Left Lane Buses and Carpools Only" are erected to indicate the use of these lanes. In addition, pavement marking on these lanes bearing "CAR POOL ONLY" or "BUS ONLY" with diamond symbols are usually used, separating them from the mixed-use lane by solid white pavement markings of 8 inches wide or more.

#### SOURCES:

- *Evaluation of Preferential Lanes for HOVs at Metered Ramps*, November 1982, Report No. FHWA-RD-82/121, Prepared by Caltrans.
- *Effect of HOV Bypass Lane Location on Violation Rates*, October 1990, Prepared by San Diego State University for Caltrans.

#### HOV Facilities

- a. *Separated facilities* - Separate roadways specifically constructed for the use of HOV providing for separation between HOV traffic and the other traffic. Usually these facilities are constructed in the median of an existing freeway with a concrete barrier. HOV can operate safely at high speeds, and the existing freeway efficiency is not reduced. Ingress/egress ramps are used when the HOV facility is physically separated from mixed-use traffic.
- b. *Reserved Freeway Lanes* - Reserved HOV freeway lanes have been used in two configurations: concurrent flow; and contraflow. Concurrent flow lanes are designated in the same direction as the peak flow on the same side of the median. This lane may be provided by adding a lane or by preempting an existing lane. Contraflow lanes are on the opposite side of the median where the

HOV traffic moves against the peak flow of traffic. The use of contraflow lanes is restricted to freeways with an imbalance in directional peak flow.

Guidelines for signing and pavement markings for HOV treatments can be found in the MUTCD.

- c. *Bus-Only Facilities* - These are either separate right-of-way facilities (e.g., Easy busway Pittsburgh, PA) or bus lanes along streets within existing right-of-way (e.g., Spring Street, Los Angeles, CA and Bellevue, WA).

#### SOURCES:

- Operational Design Guidelines for High Occupancy Vehicle lanes on Arterial Roadways, November 1994, Publication No. DOT-T-95-14.
- An assessment of High Occupancy Vehicle (HOV) Facilities in North America, Executive Report, August 1992, Publication No. DOT-T-94-17.
- High Occupancy Vehicle (HOV) Guidelines for Planning, Design, and Operations, Final Report, June 1991, Publication No. DOT-T-91-17.
- HOV Systems in a New Light, National Conference on High Occupancy Vehicle Systems, June 1994, Transportation Research Circular 442.

## 2.4.2 TECHNOLOGIES

The following technologies are used for priority access control and HOV facilities:

Separate Right-of-Way.

This is a roadway or lane developed in a separate and distinct right-of-way and designated for the exclusive use of HOVs.

Buffer and/or Concrete Barriers.

Semaphore Barrier Gates.

Semaphore barrier gates (like rail-highway grade crossing gates) come down and lock (nonbreakaway) to close HOV lane to excess traffic. They are closed at the entrance when the lanes are open in the opposing direction. Caltrans uses these gates in combination with pop-up delineators on Interstate 15. The software at the traffic management center (TMC) prevents operation the devices in an improper sequence, such as the lowering a barrier gate without the corresponding pneumatic pop-up delineators in place. The system monitors the status of all closure device sensors and prevents operation of closure devices unless the system status is known and proper.

Closed Circuit Television.

Dynamic Message Signs.

Paint Striping of Buffer for Delineation (Concurrent Flow Lane).

Bypass Lanes at Ramps (Concurrent Flow Lane).

Plastic Posts (Pop-Up) or Movable Concrete Barriers for Delineation (Contraflow Lane).

These systems consist of a barrier transfer vehicle and movable concrete barriers providing temporary physical separation between opposing flows. These barriers consist of 3-foot concrete segments joined together by pins. These barriers are installed using a transfer vehicle, and are usually transferred in the morning and afternoon peak periods to facilitate the peak traffic in each direction.

Pop-up delineator systems are operated pneumatically to raise or lower the delineators. The systems are used with detectors (most commonly used are loop detectors) to gather vehicle speeds and volumes both on the lanes and in advance of the entrance to detect gaps in the traffic.

### 2.4.3 CASE STUDIES

The San Diego Association of Governments (SANDAG) has begun a three-year demonstration project that allows 500 single-occupant vehicles to use the HOV lanes on I-15 for a monthly fee. The I-15 ExpressPass Program is initially charging \$50 per month to single-occupant vehicles that use the reversible lanes with the revenues going towards improving bus and carpool services in the area. This fee will vary and will be used to assess how pricing can be used to manage traffic and to test the value commuters place on time saving while driving.

With demand far exceeding supply, the cost has since (start of project was November 1996) increased to \$70 per month. Transponders will replace stickers in July 1997 and later electronic toll collection will be implemented where users will pay by trip and not per month. With the success of the I-15 ExpressPass, Houston is gearing towards a similar project. In this project, HOV2 vehicles will be able to use the HOV lane during the peak periods, currently restricted to HOV3 (The Urban Transportation Monitor, March 14, 1997)

The Massachusetts Highway Department has created HOV lanes on I-93, the Southeast Expressway in Boston, using the Quickchange Movable Barrier technology. It provides five inbound lanes of traffic during the morning peak period and three outbound lanes. The process is reversed for the evening peak period. The use of this technology does not require the HOV lanes to be dedicated, thus allowing the lanes to be returned to general purpose traffic after the peak traffic subsides. The 10 km HOV lane has provided 10 to 15 minute travel time savings for three-plus car poolers, vanpoolers and public transit (Traffic Technology International, August/September 1996).

Now, with the aid of reversible-lane technology, roadways in two of Dallas' busiest corridors temporarily reassign traffic flow on specific lanes, effectively doubling the number of lanes in the peak direction during morning and evening rush. In those periods, the five-lane roadways dedicate four

lanes to the peak direction-reserving one for left turns-and the fifth to the opposite direction. At other times, the roadways devote the usual two lanes to each direction, and the middle lane to left turns. Each shift takes approximately 10 minutes.

Managed from remote locations at the traffic operation center, "smart" 24-by-30-foot overhead electronic signs, placed an average of every 1,000 feet, alert motorists to the gradual switch of lane direction by displaying an amber "X" in a lane about to change direction and a red "X" once the shift is complete. Each smart sign mechanism is internally programmed with timing sequences; all report to a control center, where computers monitor and synchronize activity and alarms signal any problems. In addition to the smart signs, regular information signs on the parkways and guide markings on the pavement provide clear instructions to motorists.

With a capital expense of \$1.2 million, Dallas increased the rush-hour capacity of two older thoroughfares by 33 percent for the peak direction. The city's reversible-lane program also decreases congestion, pollution, and travel times, as it allows motorists to choose alternative routes during the ongoing reconstruction of the major freeway, scheduled for completion in the year 2000. Minimizing delays and inconvenience to commuters and businesses, the efficient and cost-effective system has enhanced access to downtown Dallas.

#### Katy Freeway (I-10 West) - Houston, Texas

The Katy Freeway HOV lane is located on I-10 West in Houston, Texas. The location of this facility,, which serves as the major travel corridor on the west side of the city. The 13-mile HOV lane was opened in stages between 1984 and 1990. It is a one-lane, barrier separated, reversible HOV lane located in the freeway median. Three park-and-ride lots and three park-and-pool lots are located in the corridor. Access and egress is provided by both slip ramps and direct access ramps. The Katy Freeway HOV lane is one of four operational HOV lanes in the Houston area and is part of a planned 96-mile HOV network.

The HOV lane is open in the inbound direction from 4:00 a.m. to 1:00 p.m. It is then closed from 1:00-2:00 p.m. to reverse the flow of HOV traffic. The lane reopens at 2:00 p.m. and operates in the outbound direction until 10:00 p.m. The vehicle occupancy requirement on the facility has changed a number of times over the life of the project. Only buses and authorized vanpools were allowed to use the facility when it opened in 1984. Due to low utilization, it was opened to authorized carpools with four or more persons in April 1985. The occupancy requirement was lowered to 3+ in December 1985, and in August 1986 it was changed to 2+ and the authorization requirement was dropped.

The 2+ occupancy requirement remained in effect until the fall of 1988. In response to the high volumes occurring in the morning peak hour, and the corresponding decline in travel speeds and travel time reliability, a 3+ vehicle occupancy requirement from 6:45-8:15 a.m. was reinstated in October 1988. The 3+ hours were slightly revised to 6:45-8:00 a.m. in May 1990, and in the fall of 1991, the 3+ requirement was applied to the afternoon peak hour from 5:00-6:00 p.m.

The vehicle volumes grew steadily after the lane was opened to 2+ carpools, reaching a high of almost 1,500 peak-hour vehicles in 1986. The vehicle and person volumes dropped initially after implementation of the 3+ occupancy requirement, but have been increasing since that time. 2 As of

December 1991, approximately 840 vehicles and 4,000 persons were using the HOV lane during the morning peak hour. In the peak period (6:00-9:30 a.m.) approximately 2,350 vehicles and 8,760 persons were using the lane.

#### I-394 - Minneapolis, Minnesota

The I-394 freeway and HOV lanes are located on the western side of the Minneapolis-St. Paul metropolitan area. The facility extends 11 miles from downtown Minneapolis to the city of Wayzata. I-394, which represents the final segment of the interstate system to be completed in the area, was constructed on the alignment of an existing arterial, US 12. Completed in the fall of 1992, the final freeway and HOV design includes two general-purpose traffic lanes in each direction and two different HOV treatments. East of Highway 100, a three-mile, two-lane, barrier-separated, reversible HOV facility is located in the median of the freeway. Those HOV lanes provide direct access into the downtown parking garages built as part of the overall project. West of Highway 100, eight miles of concurrent flow HOV lanes are in operation.

An interim HOV lane was used during construction of the I-394 facility. The interim facility was marketed as the "Sane Lane," and was implemented to help manage traffic during construction and to introduce the HOV concept in the area. The interim HOV lane was approximately three miles long, and was located in the median of US 12. Opened in November 1985, the interim HOV lane operated in the inbound direction during the morning peak period (6:00-9:00 a.m.) and in the outbound direction in the afternoon (2:00-7:00 p.m.). The operating hours changed slightly during the interim period in response to construction needs. A 2 + vehicle occupancy requirement has been in effect over the life of the project, and buses, vanpools, and carpools are allowed to use the facility.

The interim HOV lane was in operation for approximately five years. During this time, an average of some 500 vehicles carrying 1,400 persons used the facility during the morning peak hour (2). In the fall of 1992, approximately 1,100 vehicles carrying 3,580 persons were using the peak-direction concurrent flow HOV lane west of Highway 100 during the morning peak hour.

#### Route 55 - Orange County, California

Route 55 (the Newport-Costa Mesa Freeway) serves as a heavily-traveled link between the residential areas in eastern Orange and Riverside Counties and the employment centers in central Orange County. Eleven miles of HOV lanes-or commuter lanes as they are called locally-were opened on Route 55 in 1985. The Route 55 HOV facility consists of a pair of concurrent flow commuter lanes (one in each direction), and is open to buses, vanpools, and carpools on a 24-hour basis. A 2 + vehicle occupancy requirement is in effect on the Route 55 HOV lanes.

The vehicle volumes have been relatively consistent over the eight-year period, averaging between 1,100 and 1,500 vehicles during the morning peak hour in the peak direction. However, morning peak-hour vehicle volumes as high as 1,600 have been recorded on the Route 55 HOV lane. The corresponding person movements have also remained relatively constant over this period, averaging between 2,300 and 3,200 persons during the morning peak hour in the peak direction. Since very little bus service is provided in the Route 55 corridor, the vehicle volumes and person movements for the HOV lanes primarily reflect carpools.

### I-279 - Pittsburgh, Pennsylvania

The project is a four-mile, two-lane, reversible, barrier-separated HOV facility located in the median of I-279. Two short one-lane segments are located at the southern end of the facility, providing access to Three Rivers Stadium via I-579 and the downtown area via I-279. The freeway and HOV lanes were first opened in August of 1989. The HOV lanes were open to buses, vanpools, and 3+ carpools during the first three years of operation. In August 1992, a demonstration project was implemented in which the vehicle occupancy requirement on the HOV facility was lowered to two or more persons per vehicle.

The I-279 HOV lanes operate in the inbound direction from 5:00 a.m. to noon. From noon to 2:00 p.m. the lanes are closed to reverse the flow of HOV traffic. From 2:00-8:00 p.m. the lanes operate in the outbound direction with the HOV restrictions. Finally, from 8:00 p.m. to 3:00 a.m. the lanes operate in the outbound direction with no vehicle occupancy restrictions. This is done in part to accommodate traffic leaving events at Three Rivers Stadium. With the 3+ occupancy requirement, the morning peak-hour vehicle volumes had increased from approximately 164 vehicles in November 1989 to 345 vehicles in November 1991. The corresponding peak-hour person volumes had increased from some 1,100 persons to 2,200 persons. After the vehicle occupancy requirement was lowered to 2+ for a demonstration project in August 1992, the morning peak-hour volume increased to 868 vehicles and the corresponding person movement rose to 2,600.

### I-5 North - Seattle, Washington

The concurrent flow HOV lanes are located to the north of both downtown Seattle and the University of Washington. The southbound HOV lane is 7.7 miles in length and the northbound HOV lane is 6.2 miles in length. The I-5 North HOV lanes were opened in 1983 and are operated on a 24-hour basis. From 1983 until July 1991, a 3+ vehicle occupancy requirement was in effect. On July 29, 1991, the occupancy requirement was lowered to two or more persons per vehicle as part of a demonstration project.

An average of about 280 vehicles used the facility during the morning peak hour in the first few weeks following the opening of the facility. That volume had grown to 410 vehicles after the first three months of operation and 460 vehicles after the first 20 months (7, 8). Between 1985 and August 1991, an average of 460 to 550 vehicles used the HOV lane during the morning peak hour in the peak travel direction (2, 9). After initiation of the demonstration project lowering the vehicle occupancy requirement to 2+, the morning peak-hour, peak-direction volumes averaged between 1,200 and 1,400 vehicles.

Between 1985 and 1991, an average of 3,710 persons used the facility during the morning peak hour in the peak travel direction. Approximately 70 percent, or 2,605 persons, rode buses on the HOV lane, while 30 percent, or 1,105 persons, were in 3+ carpools. After the vehicle occupancy requirement was changed to 2+, the person volumes increased to an average of 5,644 during the morning peak hour in the peak travel direction. Bus ridership remained relatively constant with the reduced occupancy requirement, but the number of persons carried in carpools increased to 3,039—approximately 54 percent of the total morning peak-hour, peak-direction person volume on the facility.

### Shirley Highway (I-395) - Washington, D.C./Northern Virginia

The opening of the initial five miles of bus-only lanes on the Shirley Highway (I-395) in 1969 represented the first use of an HOV facility on a freeway in the United States. The project, which was opened in several stages between 1969 and 1975, is now approximately 11 miles in length. The two-lane, reversible HOV facility is located in the median of the freeway and is separated from the general-purpose traffic lanes by concrete barriers. Park-and-ride lots and direct access ramps are provided at strategic points along the corridor.

A number of changes have been made in the occupancy requirements and operating hours for the Shirley Highway HOV lanes. Only buses were allowed to use the facility during the first four years of operation. In December 1973, the HOV lanes were opened to vanpools and carpools with four or more persons. In January 1989, a 3+ carpool definition was implemented for the facility. Until 1985, the lanes operated in the inbound direction from 11:00 p.m. to 11:00 a.m. and in the outbound direction from 1:00-8:00 p.m. The lanes were closed for maintenance and reversing the flow of HOV traffic during other hours. As a result of a Congressionally-mandated demonstration project in the spring of 1985, the operating hours of the HOV lanes were changed to 6:00-9:00 a.m. in the inbound direction and 3:30-6:00 p.m. in the outbound direction. The lanes are open to general-purpose traffic during the remainder of the day, except when they are closed to reverse the flow of traffic. Bus service levels and service orientation were changed in 1983 with the opening of the Metrorail Yellow Line, resulting in a slight decline in vehicle and person volumes on the HOV lanes.

Approximately 39 peak-hour buses, carrying some 1,920 persons, used the HOV lanes during the first year of the project. By 1974, that number had increased to 279 buses and 11,340 passengers. As of 1991, the morning peak-hour volume for buses, vanpools, and carpools was approximately 2,773 vehicles, carrying some 18,406 persons.

## 2.5 TRANSPORTATION MANAGEMENT DURING RECONSTRUCTION

To ensure the safety of motorists and construction staff and the optimal operation of the transportation systems, it is imperative to develop a comprehensive traffic management plan for construction and maintenance work. The traffic-control plan must address key issues such as design strategies, contractual requirements, transportation policies, enforcement policies, public relations, and traffic-control strategies.

Design strategies include such measures as temporary or permanent frontage roads, narrowed lanes, and paved shoulders to provide adequate capacity during the construction period. Contractual requirements may include restriction of work hours, methods of construction and procurement, and incentive/disincentive clauses. Transportation policy issues that may be need to be addressed include: improvement to alternate routes, aggressive ride-share programs coupled with improvements to park-and-ride facilities and mass transit, and subsidized toll or transit passes. Proper enforcement of the applicable laws within the construction zone could have significant effects on the safety and operation of the transportation facility. Dissemination of the information regarding the project scope, schedule, alternative routes and modes, and addressing the public's concerns through various media (i.e., television and radio spots or news, newspaper ads and articles, the Internet, and telephone services) are keys to successful management of traffic during reconstruction projects.

Many of the traffic-management strategies discussed earlier in this and other chapters of this document (i.e., incident management, traveler information systems, and corridor management) can effectively be applied to facilities under construction. Ramp metering, ramp closures, mainline metering, HOV facilities, traffic signals and interconnect, traffic surveillance, variable lane controls, and variable speed limit are just some of the strategies that could impact the traffic operations in a construction zone.

### 2.5.1 TECHNOLOGIES

A variety of traffic-control devices has traditionally been employed in construction and work zones. The technology may vary from a static construction sign to a trailer-mounted Changeable Message Sign (CMS) to portable Highway Advisory Radio (HAR). The guideline for the design and application of the traffic control devices in a work zone is found in Part VI of the Manual on Uniform Traffic Control Devices. There are many different traffic-control devices used in construction zones and many books and manuals are dedicated to this topic. The following is not meant to be a comprehensive catalog of the available technologies, but representative of the types of devices available and their applications.

Signs.

Temporary traffic-control zone signs, convey both general and specific message. *Regulatory signs* categorized as regulatory, warning, and guide signs, inform users of traffic laws and regulations and impose legal obligations on all drivers. *Warning signs* notify drivers of general and specific conditions on or adjacent to a roadway. *Guide signs* are designed to provide drivers simple instructions and/or information about the roadway and/or route.

### Portable Changeable Message Signs (PCMS).

PCMSs are similar to dynamic message signs both in terms of functionality, as well as technology. PCMSs allow the managing agency the flexibility to display any of various messages that best suit the needs of the agency. PCMSs have a wide variety of applications in temporary traffic-control zones, including roadway or ramp closures, accident or emergency incident management, width-restriction information, advisories on road-work scheduling, traffic management and diversion, warning of adverse conditions (operationally and environmentally), and operation control.

The technologies of the message panels for the PCMS is the same as that of the DMS. However, the requirement of the control system is modified to allow field adjustment of the message, and additional requirements for the power source and mounting must be included. Unlike the DMSs, which are often controlled from Operations Centers, the PCMS messages are changed in the field. Therefore PCMSs require some method of message input and viewing capabilities. Additionally, PCMSs require portable primary and back-up power sources. The primary source could be a diesel engine or solar panels. The back-up source is often a battery system or a diesel engine. The sign is either mounted on a trailer pulled by a separate vehicle or mounted on a truck.

### Arrow Displays.

Arrow displays may be used in the arrow or chevron modes for stationary or moving lane closures. The display is rectangular in shape with a nonreflective black finish. The color of the limit-emitting elements is yellow and can be produced by similar technologies, as described under the DMS section.

### Channelizing Devices.

The function of channelizing devices is to warn and alert drivers of conditions created by work activities in or near the traveled way, to protect workers in the temporary traffic-control zone, and to guide motorists and pedestrians safely. They should provide for a gradual shift of traffic from one lane to another or onto a bypass or detour, and/or accommodate a reduction in traveled way. They may also be used to separate mixes of traffic (i.e., single-occupancy vehicles, HOV, trucks, pedestrians, bicycles) or opposing traffic. Channelizing devices include but are not limited to cones, tubular markers, vertical panels, drums, barricades, temporary raised islands, and barriers.

### Pavement Markings.

Similar to channelizing devices, pavement markings are a part of a system of devices used to manage the traffic in construction zones. They must be complemented by other devices to effectively guide the motorist through a construction zone. Pavement markings may be categorized as striping, Raised Pavement Markers (RPM), and delineators.

### Portable Barriers.

Portable barriers are designed to prevent vehicles from penetrating work areas while minimizing vehicle-occupant injuries. The portable barriers have been in use for many years and have proven to be an effective method of separating travel way from the work area. They may also be used for channelization purposes. Because of their portability, these barriers can easily be moved by time of

day or in case of an incident to accommodate for additional lanes and capacity in the peak direction. Portable barriers may be constructed of concrete, metal, polymerized fiber (filled with water or sand), or any material that can physically prevent vehicular penetration.

#### Temporary Traffic Signals.

Temporary traffic signals can be used for special applications in temporary and long-term work areas. Regardless of their temporary nature, they must meet and maintain the standards set forth in the MUTCD. The design and functionality of these signals is not much different from those of the permanent signals except for the use of such techniques as alternate power sources and portable loops or video detection.

#### Portable Highway Advisory Radio (HAR).

Portable HARs effectively provide construction and traffic-related information to the motorist in advance of or within a construction zone. The technology remains the same as the permanent HAR systems except for the alternate power source and mobility requirements. There are trailer and van-mounted HAR systems in use today. This technology could be very beneficial since it could serve a large audience.

## 2.5.2 CASE STUDIES

Smart Work Zone technology uses cameras to monitor work areas in the Minneapolis-St. Paul area and relays the information to the local traffic center, which in turn uses DMSs to alert motorists about work-area caution zones. A survey by MnDOT revealed that the public is getting the message. Most of those polled said the warning signs were accurate, useful, and gave them the information they needed. In addition, MnDOT posts the work-zone information on the Internet (ITS World, Jan/Feb 1997).

#### SOURCES:

- *Manual on Uniform Traffic Control Devices (MUTCD)*, September 1993, Publication No. FHWA-HI-94-027.
- *Special Report 212 - Transportation Management for Major Highway Reconstruction*, 1996, Transportation Research Board.
- *Proceedings of the Symposium on Work Zone Traffic Control*, June 1991, Publication No. FHWA-TS-91-003, Hugh W. McGee, Lynne F. McGee, Nancy L. Geisler .
- *Corridor Transportation Management for Highway Reconstruction*, May 1986, Publication No. DOT-I-86-35, William T. Steffen, Sonia Weinstock, Mary Ellen Sullivan.
- *Corridor Traffic Management for Major Highway Reconstruction - A Compilation of Case Studies*, September 1986, Publication No. HPN-23/R11-86(300).



## 2.6 DETECTION

Electronic freeway detection systems provide real-time traffic information on volumes, speed, and occupancy of each lane of the freeway. Congestion is identified through algorithms that use data from constantly monitoring detectors. With accurate information, traffic management agencies can implement control strategies, detect and manage traffic during incident conditions, and activate motorist information systems that will improve the quality of operations in the traffic network.

### 2.6.1 INSTALLING DETECTION SYSTEMS

A communication system linking the central control room with detector systems, video cameras (CCTV equipment, video imaging, etc.), and changeable-message-sign controllers is essential to the optimal operation of detection and control systems. The Freeway Management Handbook (DTFH61-95-C-00128) discusses the process for planning, designing, and installing detection, as well as video surveillance systems. This process a system engineering approach that is an iterative process whereby operational problems are identified, system concepts and objectives are established, potential solutions are developed and evaluated, new solutions are identified, and system objectives are redefined. This approach can be used for: planning and implementing a new system where no current transportation management system exists; modifying an existing system to add new transportation management functions; or upgrading the technology in an existing traffic management system to current standards. This process is followed in this section, Detection, to illustrate how it can be used for planning a system. It can also be used for other strategies/technologies throughout this document. The following steps make up this process.

#### Problem Identification

The first step in the process is to identify the problems to be addressed by the detection system. Issues to be addressed in the problem identification stage are:

- Identifying and locating operational problems;
  - Determining functions to be performed by the detection system; and
  - Inventorying existing detection capabilities.
- a. *Identify and Locate Operational Problems.* This process involves identifying those freeways that would most benefit from the use of detection systems: those with significant amounts of congestion. These areas may be ones where significant increases in traffic demand are expected, where significant amounts of maintenance or construction activities occur, where there is increased delay, where there are slower and inconsistent travel speeds, and where there is a high frequency of traffic incidents.
  - b. *Determine Functions of Detection System.* A detection system may serve several objectives, including:
    - Detection of incidents that impact traffic operations

- Monitoring traffic operation and supporting the implementation of control strategies (e.g., lane control, ramp metering, etc.);
  - Monitoring environmental and pavement conditions; and
  - Monitoring traffic during construction and/or special events.
- c. *Importance and Types of Data.* The importance and type of data to be collected must be identified for selection of the appropriate detection system. Both real-time and historical data may be needed for traffic management, to monitor current traffic operations as well as to establish a record of past traffic conditions. The types of data to be collected may include the following.
- Speed
  - Volume
  - Occupancy
  - Vehicle travel times
  - Bus location
  - Queue length
  - Pavement condition

In addition, the accuracy and cost of the detection system must be balanced. The accuracy of the data may be more important for some systems (e.g., incident detection). The accuracy of the data as well as the traffic parameter range and collection interval are presented in Table 2.6.1

d. *Inventory Existing Detection Capabilities.*

The existing detection resources and communications system must be identified and evaluated to determine their suitability for continued use. It is important to determine if these systems meet existing needs and if they can accommodate changes in system requirements. It is also very important to evaluate the communications system for the various detection technologies and to identify the communications requirements.

#### Identification of Partners

Another important step in the implementation of a successful detection system is to identify the partners that are to be involved. These partners come from three areas: intra-agency; interagency, and other.

- a. *Intra-agency.* The intra-agency partners may be representatives from the following areas: management; planning; design; operations; and maintenance. This team of individuals is identified from the start of the planning stage and is crucial in the decision-making process. It is essential to have commitment from top management as well as support staff to assure so that resources are allocated to these systems and to assure system success.

- b. *Interagency.* The interagency partners may be State, city, and/or county representatives, to ensure that proper coordination is maintained between jurisdictions. This coordination will help in identifying problems and minimizing the effects of certain activities and events on overall traffic operations. It is therefore important to exchange real-time information between these jurisdictions on a continual basis, and to maintain a database for use by such entities as law enforcement, emergency, media, wreckers, and other information services, as well as commercial vehicle operators and the general public. (Appendix C presents an interagency agreement for a Regional Transportation Management System in Harris County, Texas).
- c. *Additional Resources.* Other groups of resources to be considered for identifying and evaluating the various detection technologies include: manufacturers and suppliers, the users, the researchers, consultants and other interested groups. Information gathered from these groups will help in the identification and determination of the state-of-the-art as well as the state-of-the-practice, and will help in bridging the gap between the two.

### Establish Goals and Objectives

Establishment of the goals and objectives of a system, requires identification of what the system is to accomplish. Goals are used to define the system's long-range functions. Objectives define the level of performance to be expected in the future.

The detection system provides support for other elements of a transportation management system (e.g., incident management, traveler information, ramp control, etc.). The goals and objectives of the detection system, therefore, relate to the goals and objectives of the elements supported by it. For example, the goals of an incident management system may be to reduce the impact of incidents on traffic operations. The objectives of the system may be to detect an incident in less than two minutes and reduce the incident clearance time by five minutes. The detection system can, therefore, be evaluated by determining the system's ability to meet these established objectives.

### Establish Performance Criteria

There is a wide range of traffic detectors from which to choose, and with advancements in technology, the alternatives are becoming even greater. It is therefore important to develop performance criteria to help in the selection process. The performance criteria should be related to the ability of the system to meet the established goals and objectives. Criteria that may be used to measure the performance of a detection system include: reliability of the system (e.g., percent of time the system is producing desirable results, amount of maintenance required); accuracy of the data (e.g., percent of false alarms, difference between measured data and actual data); timeliness of the data (e.g., time between detection and actual occurrence). These criteria can be used to establish parameters to evaluate detection systems.

### Define Functional Requirements

The next step in the decision-making process involves defining all of the functions necessary to achieve the established system objectives. The functions should be defined independent of the available technologies. These functional requirements of the detection system should be dependent

upon the element that it will support (e.g., elements that use such data as speed, volume, occupancy, travel time, queue length, headway, etc.).

### Define System Architecture

After defining the functional requirements of a detection system, the next step is to define the system architecture. This is the framework within which the system carries out the functions required to support the established objectives. Information concerning the development of system architecture for transportation management systems has been prepared through the national architecture development effort for ITS. The system architecture for the detection system should be based on this national development and defined to meet local issues and concerns.

### Identify and Screen Technologies

Once the functional requirements and architecture of the detection system have been defined, the next step is to identify those technologies that meet the defined requirements. The steps in identifying and choosing the appropriate technologies include the following.

#### a. *Identification of Alternative Technologies.*

There are a number of technologies available for collecting traffic data. The characteristics, applications, and requirements for various detection technologies are discussed in the following section. Available detection systems and their capabilities are constantly changing due to technological advancements and system improvements. It is therefore important to identify the state-of-the-art in detection systems. It is also important to continually interface with the following groups and individuals during the identification process: manufacturers, users, consultants, researchers, other interested individuals. Numerous sources of information (see below) are available for assistance in identifying available and evolving detection technologies. In addition, the Internet has a wealth of information on various manufacturers of detection systems, as well as sites for information on detection technologies. (see below)

#### b. *Evaluation of Alternative Technologies.*

When evaluating detection technologies, the following factors should be considered for each detector: location of detector (i.e., embedded in pavement or nonintrusive); installation, operation, and maintenance requirements; reliability; expected life; life-cycle costs; type of communication medium available; requirements for future expansion. After the initial screening, which will look at these factors, the detailed evaluation should include the following: estimating costs and benefits of each alternative; comparative analysis; and selecting the system offering the most potential.

Numerous extensive and comprehensive studies have been performed to test and evaluate detection technologies. The two most recent are the Hughes Aircraft study ( ) and the Minnesota Guidestar study ( ).

In the Hughes Aircraft study, off-the-shelf, state-of-the-art detectors obtained from various manufacturers were field tested and evaluated. The detector technologies are shown in Table 2.6.1 and the site locations, test period and the amount of data collected are shown in Table 2.6.2. Table 2.6.3 provides a qualitative assessment of the detector technologies tested under various traffic parameters.

Table 2.6.1

Symbol	Technology	Manufacturer	Model
U-1	Ultrasonic Doppler	Sumitomo	SDU-200 (RDU-101)
U-2	Ultrasonic Presence	Sumitomo	SDU-300
U-3	Ultrasonic Presence	Microwave Sensors	TC-30C
M-1	Microwave Radar Motion Medium Beamwidth	Microwave Sensors	TC-20
M-2	Microwave Radar Doppler Medium Beamwidth	Microwave Sensors	TC-26
M-4	Microwave Radar Doppler Narrow Beamwidth	Whelen	TDN-30
M-5	Microwave Radar Doppler Wide Beamwidth	Whelen	TDW-10
M-6	Microwave Radar Presence Narrow Beamwidth	Electronic Integrated Systems	RTMS-X1
IR-1	Active IR Laser Radar	Schwartz Electro-Optics	780D1000(Autosence 1)
IR-2	Passive IR Presence	Eltec	842
IR-3	Passive IR Pulse Output	Eltec	833
IR-4	Imaging IR	Grumman	Cat Eye
VIP-1	Video Image Processor	Econolite	Autoscope 2003
VIP-2	Video Image Processor	Computer Recognition Systems	Traffic Analysis System
VIP-3	Video Image Processor	Golden River Traffic	Marksman C-CATS 810
VIP-4	Video Image Processor	Sumitomo	IDET-100
VIP-5	Video Image Processor	EVA	2000
A-1	Passive Acoustic Array	AT&T	SmartSonic TSS-1
MA-1	Magnetometer	Midian Electronics	Self-Powered Vehicle Detector
L-1	Microloops	3M	701
T-1	Tube-Type Counter	Timemark	Delta 1

Table 2.6.2

<b>Location</b>	<b>Test Period and weather</b>	<b>Runs</b>	<b>Data Collected (MB)</b>	<b>Traffic-Flow Direction</b>
Minneapolis Freeway I-394 at Penn Ave.	Winter 1993; Cold, snow, sleet, fog	15	200	Departing (am) and Approaching (pm)
Minneapolis Surface Street Olson Hwy, at Lyndale Ave.	Winter 1993; Cold, snow, sleet, fog	7	32	Approaching
Orlando Freeway I-4 at SR 436	Summer 1993; Hot, humid, heavy rain, lightning	28	670	Approaching
Orlando Surface Street SR 436 at I-4	Summer 1993; Hot, humid, heavy rain, lightning	21	200	Departing
Phoenix Freeway I-10 at 13th Street	Autumn 1993; Warm, rain	32	868	Approaching
Tucson Surface Street Oracle Rd. at Auto Mall Dr.	Autumn 1994; Warm	34	2892	Departing
Phoenix Freeway I-10 at 13th Street	Summer 1994; Hot, low humidity, some rain	31	1060	Approaching

Table 2.6.3

<b>Technology</b>	<b>Low-Volume Count</b>	<b>High-Volume Count</b>	<b>Low-Volume Speed</b>	<b>High-Volume Speed</b>	<b>Best in Inclement Weather</b>
Ultrasonic	-	-	-	-	-
Microwave Doppler	x	x	x	x	x
Microwave True Presence	x	x			x
Passive Infrared	-	-	-	-	-
Active Infrared	-	-	-	-	-
Visible VIP	x	x			-
Infrared VIP					
Acoustic Array	-	-			
SPVD Magnetometer	x	-	-	-	x
Inductive Loop	x	x	-	-	x

In the Minnesota Guidestar study, nonintrusive detection technologies were field tested and evaluated. In the initial equipment field test, the nonintrusive detectors shown in

Table 2.6.4 were tested at Interstate 394 and Penn Avenue in Minneapolis. In the extended field test, the detectors shown in Table 2.6.5 were tested and evaluated at the same site as the initial field test. The extended field test evaluated detector technologies under a variety of environmental and traffic conditions. Appendix D provides a summary of the findings with Figures.

Table 2.6.4 Devices Evaluated in the Initial Field Tests

<b>Technology</b>	<b>Vendor</b>	<b>Device</b>
Passive Infrared	Eltec Instruments	Model 842 Sensor
Passive Magnetic	Safetran Traffic Systems	232E IVHS Sensor
Radar	Electronic Integrated Systems	RTMS
Doppler Microwave	Microwave Sensors	TC 26B
Doppler Microwave	PEEK Traffic	PODD
Doppler Microwave	Whelen Engineering	TDW 10
Doppler Microwave	Whelen Engineering	TDN 30
Passive Acoustic	International Road Dynamics	SmartSonic
Pulse Ultrasonic	Microwave Sensors	TC 30C
Video	Rockwell International	TraffiCam

Table 2.6.5 Devices Evaluated in the Extended Field Tests

<b>Technology</b>	<b>Vendor</b>	<b>Device</b>
Passive Infrared	Eltec Instruments	Model 833 Sensor
Passive Infrared	ASIM Engineering, Ltd.	2-IR-224
Active Infrared	Schwartz Electro-Optics	Autosense I
Passive Magnetic	Safetran Traffic Systems	2-232E Sensor
Radar	Electronic Integrated Systems	RTMS
Doppler Microwave	PEEK Traffic	PODD
Doppler Microwave	Whelen Engineering	TDN 30
Passive Acoustic	International Road Dynamics	2-SmartSonic
Pulse Ultrasonic	Microwave Sensors	TC 30C
Pulse Ultrasonic	Novax	Lane King
Video	Rockwell International	TraffiCam
Video	Image Sensing Systems	Autoscope 2004
Video	ELIOP Trafico S.A.	EVA 2000 S
Video	Peek Transyt	Video-Trak 900

c. *Selection of Appropriate Technology.*

The conclusion that the best system is the one with the greatest benefit:cost ratio is not always correct. For example, a simple, low-cost system with fewer benefits may have the same benefit-cost ratio as a more sophisticated, affordable system with more benefits. Therefore, the analysis should include allowable expenditure for the system and the net benefits of each alternative.

#### Plan Development

After the system's detection technologies have been selected, the next step is to develop a plan for implementation. The purpose of the Implementation Plan is to ensure that the system is designed, built, operated, and maintained to accomplish its purpose in the most efficient manner possible. The Implementation Plan is required when either a traffic control system or an expansion of an existing system is Federally funded. The Implementation Plan documents the results of the previous steps and identifies how the system will be implemented in the field. The Implementation Plan should also assess the phasing, procurement, and funding options for implementing the system. For further discussion on the contents of the Implementation Plan, see Appendix E. The operations and maintenance requirements of the planned systems must be in balance with the availability of proper personnel, equipment, and budget resources to operate and maintain the systems. More information of operations and maintenance of systems is provided in Appendix F.

#### Identify Funding Sources

Funding sources should be identified during development of the Implementation Plan. Funding for detection systems may come from both the Federal and State sources. On the Federal level, funding is available through ISTEA. Under ISTEA, Federal aid is available to assist with capital and operating expenses for transportation management programs. By including projects in the Transportation Improvement Program (TIP), a spending plan required by the Federal government, agencies can receive ISTEA funds. ISTEA permits startup and operating costs from the following sources: Surface Transportation Program (STP); National Highway System (NHS); and Congestion Mitigation and Air Quality Program (CMAQ). ISTEA funds are generally used for system deployment and to provide startup assistance. Public/private partnership is another approach that many agencies are using to increase the source of funding for transportation management systems. Further discussion and details of funding opportunities for transportation management systems at the local, State, and Federal levels is found in Chapter 9.

#### Implementation and Deployment

The implementation process includes the activities involved in installing the components of the detection system to meet established goals and objectives. Issues to be considered in the implementation of detector systems include: phasing of installation or construction; training of operators and maintenance personnel; and system deployment approaches. The procurement strategies include: sole-source; engineer/contractor; two-step approach; system management; design-build. Appendix E discusses these strategies.

## Evaluation

Once a detection system has been installed, it is important to evaluate its effectiveness in meeting the objectives. Numerous methods of evaluating the effectiveness of detection systems and other transportation management systems are used by practitioners. The most common method is the before-and-after method. By this method, the performance of the transportation management system is measured before the system is implemented and compared with measures of the same performance criteria after implementation. Appendix G provides other evaluation procedures for selected ITS strategies.

## 2.6.2 TECHNIQUES/STRATEGIES

### Vehicle Detectors

Vehicle detection systems identify unforeseen reductions in freeway capacity and allow initiation of appropriate procedures to restore the freeway to full capacity. The inductive-loop detectors are the most extensively used today.

### Call Boxes

Call boxes are often installed along freeways to provide motorists a means of calling for assistance. They are also used for incident detection, to inform the operating agency of the incident and its nature. Call boxes have become multitasking—not only for motorist aid but also for the acquisition and transmission of traffic data and weather and environmental data. Other capabilities include: full duplex voice capability; transmission and receipt of digital data; and detection and processing of real-time traffic information.

### CB Monitoring

Drivers of CB radio-equipped vehicles report observed incidents to a central control center via channel 9, which was designed as an official emergency channel by the Federal Communications Commission (FCC) in 1970.

### Weather and Environmental Detection

These systems consist of: surface sensors, which monitor pavement temperature and surface conditions including presence of ice, frost, water, and snow; atmospheric-condition sensors, which monitor air temperature, dew point, relative humidity, precipitation, wind direction, and wind speed; remote processing units, which collect and transmit the surface and atmospheric data from the sensors to a central processing unit; and central processing unit, which processes all the data for graphic presentation and transmits the data to remote terminals. During the last 20 years, a number of State highway agencies have installed ice-detection and weather-information systems. Evaluations have included the performance of the system equipment as well as its usefulness, effects on highway safety, and cost-saving aspects. To encourage and assist State Highway agencies in evaluating their systems, the FHWA initiated Test and Evaluation Project No. 11, Ice Detection and Highway Weather Information Systems, in 1988. The project was originally identified as Special Experimental Project No. 13, but

the name was changed in 1990. Eight cooperating agencies agreed to evaluate their use of these systems during the winters of 1989 and 1990. All participants used the SCAN system and SCAN-CAST service from Surface Systems, Inc. Appendix H provides a summary of this effort.

### Overheight Vehicle Detection

These detectors detect the presence of overheight vehicles to prevent damage to the vehicles and highway structures. The detection and warning systems use infrared light, with a transmitter of the light on one side of the highway and a receiver on the other side. When an overheight vehicle is detected, the driver is warned of an approaching structure by means of a system of alarm bells and flashing lights. These systems are used on overpasses, approaches to tunnels, approaches to parking garages, etc. The detectors are used in conjunction with warning message devices including blankout signs, static signs with flashers, and audible warning devices.

### Automatic Truck Warning System

Truck rollover on freeway interchange ramps is a serious safety problem. Large trucks overturn or rollover on curved ramps as the truck's driver attempts to negotiate the curve. Severe traffic disruption often results from these rollovers and the severity can be multiplied by the type of cargo the truck is hauling. In concept, a truck warning system must perform three primary functions: collection of data about the vehicle and highway, determination of rollover risk, and operation of a device to warn the truck's driver of the risk with sufficient time to take corrective action. It is possible to design a warning system that is entirely infrastructure-based or entirely vehicle-based. In addition, it is conceivable to design an integrated system that uses both infrastructure information and data derived from on-vehicle measurements.

## 2.6.3 CASE STUDIES

The following are deployments and/or operational tests that apply state-of-the-practice and advanced technologies in automatic truck warning systems:

### DC Beltway Demonstration Project

This demonstration project, installed at three locations on the Washington, DC, Beltway, includes the state-of-the-practice for an infrastructure-only automatic truck warning system. Piezoelectric, inductive loops and optical sensors measure the weight, speed, and height of passing vehicles as they approach the freeway ramp, and based on that data, trucks are identified and classified. If the estimated lateral acceleration is higher than the preset rollover threshold for the identified class and type of vehicle, warning lights on a roadside sign are activated until the vehicle passes the sign.

### NHTSA Cooperative Agreement

A vehicle-based rollover system is in development under this cooperative agreement whereby the roll behavior of a tractor-trailer combination is measured by sensing the forces on the fifth-wheel coupler between tractor and trailer. The measured forces are then used with a computerized model of the vehicle suspension to estimate how closely the vehicle approaches rollover. The estimated risk is to

be displayed on an in-cab display, allowing the driver to learn how the vehicle behaves and avoid high-risk driving. This development is for training purposes only.

#### SOURCES:

- *Feasibility of an Automatic Truck Warning System*, September 1993, Publication No. FHWA-RD-93-039.
- *An Automatic Warning System to Prevent Truck Rollover on Curved Ramps*, Public Roads, Spring 1994, Vol. 57, No. 4, by Hugh W. McGee and Rodney R. Strickland.
- *Invehicle Safety Advisory and Warning System (IVSAWS)*, Vol. II: Final report, March 1996, Publication No. FHWA-RD-94-190, Prepared for FHWA by Hughes Aircraft Company.
- *Heavy Vehicle Intelligent Dynamic Enhancement Systems*, NHTSA Cooperative Agreement, Publication No. DTNH22-95-H-07002.

## 2.6.4 TECHNOLOGIES

### Vehicle/User Devices

Technologies for vehicle/user devices are described below.

#### a. *In-Road Detection/Warning Systems*

In-road detection/warning systems use sensors deployed in or alongside the roadway just before a hazardous off-ramp. These sensors detect trucks and measure characteristics of the vehicles. Based on those measurements and the known lateral acceleration demand of the upcoming off-ramp, a roadside CPU estimates the risk of rollover for that vehicle. If that estimated risk exceeds a threshold set by the responsible local agency, the central processing unit (CPU) activates flashing lights or another warning device to alert the truck's driver to the rollover hazard.

Various sensors (e.g., piezoelectric, inductive loop, optical, video imaging) are used to measure the weight, speed, and height of trucks as they approach freeway off-ramps. Based on this data, trucks are identified and classified by a local controller (CPU) and separated on the basis of height (i.e., greater or less than 11 feet high) into box and tanker. Speed is measured at two points, allowing longitudinal acceleration to be estimated, and speed during the turn on the off-ramp, is projected with estimation lateral acceleration in the turn. If the lateral acceleration, that is estimated, is higher than a preset rollover threshold for the type and class of truck identified, warning signs and/or signals (e.g., sign with or without some type of beacon) on the roadside are activated until the truck passes.

Overheight detectors (see above) are used to determine if a truck's height is above a threshold height. The weight of the truck is usually obtained using commercial Weigh-In-Motion (WIM) systems. These systems use a combination of inductive-loop and piezo sensors. A controller is needed to accept all of the electrical inputs from the detection devices, process the information, and send signals to activate the warning devices. This controller is usually customized (i.e., no controller exists

off-the-shelf to accept all of the inputs from all detectors and process all information) to accept and process the information. Warning devices usually consist of two types: a static sign with yellow beacons that flash when activated, and a static sign in combination with a dynamic message sign (DMS) (e.g., blankout sign) that displays a message such as "Trucks Reduce Speed" when activated.

b. *In-Vehicle Detection/Warning Systems*

The in-vehicle detection/warning system measures the actual roll behavior of a truck under lateral acceleration and alerts the driver through an in-cab signal of the rollover danger. The system consists of equipment installed both in the vehicle and at roadside. The components of the systems follow.

1. *Roadside Transponder* - Transmits information on ramp characteristics such as ramp radius, superelevation, etc.
2. *Roadside Antenna* - Relays the information from the roadside transponder to the truck
3. *Roadside Detectors* - Detects trucks and their speed, weight, and activates the transponder when the vehicle approaches the ramp.
4. *In-Vehicle Electronic Reader* - Receiving and reads ramp information sent by the antenna.
5. *In-Vehicle Computer* - Processes all driver-input (e.g., vehicle, cargo, load type, vehicle height), and roadside data, and activates an in-vehicle warning device if the rollover threshold has been exceeded.

There are many technologies available for collecting traffic data. They range from the most used, inductive loop, to the newest and most promising technology, video image processing.

Vehicle Detection

The following are available technologies used for vehicle detection.

a. *Inductive Loop*

The inductive-loop detectors are the most extensively utilized today. They provide data on volume, speed, and density (occupancy). The technology consists of a loop wire of one or more turns embedded in the pavement and connected to an electronic amplifier (located in the controller cabinet). This detector identifies the presence or passage of a vehicle.

Inductive loop systems can provide data on volume, speed, occupancy, presence, headway, and classification.

Advantages:

- Flexible design.
- Wide range of applications.
- Provides basic traffic parameters.

Disadvantages:

- Installation requires pavement modifications.
- Installation and maintenance require lane closures.
- Detectors subject to of traffic stresses.
- Subjection to damage and dislocation by road traffic or roadway equipment such as sweepers and snow plows.

b. *Magnetometer*

This small, cylinder-shaped detector is placed in a drilled hole in the road for limited use in real-time surveillance. Limited applications include use under bridge decks, when damage to the reinforcing steel becomes a concern, and in roadways where deteriorating pavement dictates minimal saw cut.

Advantages:

- Easier to install than loops.
- Can be used in situations where loops are not feasible (e.g., bridge decks).
- Less susceptible to traffic stresses than loops.

Disadvantages:

- Installation requires pavement modifications.
- Installation and maintenance require lane closures.
- Small detection zone.
- Typically used to provide count and occupancy only.

c. *Radar/Microwave*

This detector is non-pavement intrusive and is commonly used to monitor vehicle speeds for law enforcement and traffic management applications. Some advanced radar/microwave detectors can also be used as presence detectors. Because they use electromagnetic energy, radar/microwave sensors are typically unaffected by weather conditions, especially when measurements are made from a short distance. They can also be used for both day and night operations. One problem with radar/microwave detectors is that they tend to lock onto the strongest return signal, which is often produced by the largest vehicle in the traffic stream (a large truck). Smaller vehicles may not be detected by some units. Therefore, the vehicle mix of the traffic stream may affect the accuracy of the data supplied by this type of technology.

Radar/microwave systems can provide volume, speed, occupancy, and presence detection.

Advantages:

- Generally insensitive to weather conditions.
- Provides day and night operation.

Disadvantages:

- Requires FCC license for operation and maintenance.
- May lock on to the strongest signal (e.g., large truck).
- Susceptible to interference from other devices operating at the same frequency.

d. *Ultrasonic*

These detectors (e.g., smartsonic) are passive acoustic sensors and automated-signal and information-processing systems that listen (no energy is radiated by the system) for the noise generated by stationary or moving vehicles in a detection zone on the roadway. Only those vehicle sounds from within a specific detection zone are retained. Sounds from locations outside the detection zone (such as an adjacent lane) are severely attenuated and are ignored. The detection process for vehicle sound energy is analogous to the way the metal in vehicles is detected as those vehicles pass over a loop. It is an overhead-mounted system with limited side mount. Ultrasonic detection fully emulates loop-output signals, and thus requiring no modification to existing system hardware or software.

Ultrasonic systems can provide volume, speed, occupancy, presence, and classification data.

Advantages:

- Completely passive
- Generally insensitive to weather conditions
- Provides day and night operation

Disadvantages:

- Relatively new technology for traffic surveillance
- Environmental conditions such as winds, heavy snowfall, and rain may inhibit the propagation of sound waves.

e. *Infrared*

There are two types of infrared detectors: active and passive. *Active infrared detectors* focus a narrow beam of energy onto an infrared-sensitive cell, and vehicles are detected when they pass through the beam, interrupting the signal. These detectors can be used either as presence or pulse detectors. Detector performance can be affected by weather conditions (fog, rain, snow) causing inconsistent beam patterns. *Passive infrared detectors* do not transmit energy, but measure the amount of energy emitted by objects in their field of view.

Infrared detection systems can provide volume, speed, occupancy, presence, and classification data. Active infrared detectors depend on the reflection of transmitted laser beams.

Advantages:

- Active detector emits narrow beam for accurate determination of vehicle position.
- Provides day and night operation.
- Provides most basic traffic parameters.
- Passive detectors can be used for strategic loop replacement

Disadvantages:

- Operation affected by precipitation such as rain or fog.
- Difficulty in maintaining alignment on vibrating structures.
- Susceptible to sudden changes in background radiation due to rain or clouds.
- Some surfaces such as windshields and black metal and plastic car bodies are poor reflectors.

f. *Video Image Processing*

With video image processing, cameras provide images that are used by a video processor to emulate traffic data. It is possible to define multiple detection locations within the camera viewing area. These "pseudo-detectors" are not fixed, but may be moved by the operator if desired. The type of signal processing algorithm used by the image processor dictates the type of data obtainable by the system.

These systems can provide volume, occupancy, and presence detection. In more advanced systems, individual vehicles are tracked as they pass through the field of view, allowing identification of speed, vehicle classifications and travel times in the detection zone. Most processing algorithms have been optimized counter shadows, illumination changes, and reflections.

Advantages:

- Location or addition of detector zones can easily be done on the PC.
- Provides basic traffic parameters.
- Provides wide-area detection.

Disadvantages:

- Inclement weather, shadows, and poor lighting can affect performance.
- May require significant processing power and a wide communication bandwidth.

Table 2.6.6 provides a list of detector technologies with their functional capabilities and cost. Table 2.6.7 provides estimated operating and maintenance costs for various detection technologies. Table 2.6.8 Summarizes the detector technologies discussed and the applications for which they are best suited.

Table 2.6.6 Characteristics of Traffic Detectors

Detector	Applications						Lane Coverage per Sensor	Bandwidth	Life	Reliability	Technology	Cost	
	Count	Presence	Speed	Occupancy	Classification	Installation						Detector (each)	
Inductive loop	X	X	X	X	X		Size of loop	Low	Moderate	Current	Moderate (\$1000)	Low (\$500-800)	
Magnetometer	X	X	(1)	X			Single lane	Low	High	Current	Moderate (\$1000)	Low-Moderate (\$500-1,500)	
Microwave radar	X	(2)	X	(2)	(2)		Multiple	Moderate	High	Current	Low (\$500)	Low-Moderate (\$700-3,000)	
Infrared	X	X	X	X	(2)		Single (active); Multiple (passive)	Low-Moderate	High	Developing	Low (\$500)	Moderate-High (\$1,000-8,000)	
Ultrasonic	X	X	X	X	(2)		Single	Low	High	Developing	Low (\$500)	Low-Moderate (\$600-1,500)	
Acoustic	X	X	X	X	(2)		Multiple	Low-Moderate	N/A	Developing	Low (\$500)	Moderate (\$1,500)	
Video Image Processing	X	X	X	X	X		Multiple	Moderate-High	Moderate	Developing	Low (\$500)	Very High (\$10,000-25,000)	

Table 2.6.7 Estimated Operating and Maintenance Costs for Detection Systems

Description	Base Unit	Estimated Annual Operations Cost/Unit	Estimated Annual Maintenance Cost/Unit	Estimated Annual Combined O&M Cost/Unit	Cost Assumptions
Inductive Loop	Per station	\$0	\$200 - \$300	\$200 - \$300	Costs include contract maintenance/replacement of loops. Costs assume four lanes per station, with two loops per lane. Costs also assume loop failure rates of 4% to 6% per year.
Radar, Ultrasonic, Infrared	Per station	\$0	\$200 - \$300	\$200 - \$300	Costs include routine maintenance of detectors. Costs assume four lanes per station, with one detector per lane.
Video Image Processing	Per station	\$0	\$500	\$500	Costs include routine maintenance and calibration of detectors. Costs assume one video detector per station.

Table 2.6.8 Traffic Detector Applications

Monitoring Application	Inductive Loop	Magnetometer	Microwave/Radar	Infrared		Passive	Ultrasonic	Video Image Processing
				Acoustic	Active			
Incident Detection	X		X					X
Traffic Conditions	X	X	X	X	X	X	X	X
Special Event	X		X	X	X	X	X	X
Implement Control Strategies	X	X	X	X	X	X	X	X
Vehicle Classification	X				X			X
Ramp Metering	X							X

## 2.6.5 CASE STUDIES

TransGuide - From the TransGuide ITS Design Report ([transguidewwwserver.datasys.swri.edu:80/articles/designreport.html](http://transguidewwwserver.datasys.swri.edu:80/articles/designreport.html) )

### Loop Detectors

Loop detectors were chosen as the method of detecting incidents based on several factors. One significant factor is the climate in San Antonio, which is relatively amenable to buried detection technology. There are essentially no extended periods of time when the ground freezes. There is also very little snow and therefore no salt is used on the roadways. An additional factor is that loops were already included in some of the existing freeways when they were constructed. Finally, loop detection technology is a well understood and mature detection technology. While other detection technologies may be applied in the future, the initial sections of the TransGuide ITS are being implemented using loop detectors.

Other detection technologies available include radar detection, visual or video detection, and audio based detection. Each of these technologies is being used in experimental or operational systems. However, none is as mature as the loop detector and none was demonstrably better or cheaper than loop detector technology when the TransGuide system was specified.

While loop detectors may be susceptible to reliability problems in some locations, they are well suited to the San Antonio climate. Radar detectors are being experimented with in other cities; however, they do not offer specific information on separate vehicles. Radar detectors also require overhead installation, which may increase installation and maintenance costs. Video detection methods are beginning to be available, but they are relatively expensive and may suffer from some accuracy problems in specific vehicle or environmental conditions. Audio detection methods are also being developed, but they don't offer specific information on every vehicle and lack a performance history.

Loop detectors are buried in pairs approximately every ½ mile along the main lanes of the freeway, although freeway geometry causes this spacing to be less in some sections. Isolated loop detectors are also buried at strategically located positions, such as on and off ramps. The single loop detectors can determine the count of vehicles passing over the detectors (volume) and the percentage of time a vehicle is over the sensor (occupancy). In addition to volume and occupancy, pairs of loop detectors can be used to determine the average speed of vehicles passing over the loops.

The loop detectors consist of six foot by six foot loops buried one inch under the roadway. The loops are constructed using one 14-American Wire Gage (AWG) conductor. Loop pairs are installed 12 feet apart and are made up of differing numbers of turns to minimize crosstalk. Changes in a loop's inductance are detected to sense the presence of a vehicle.

The loop detector signals are analyzed by LCUs similar to those used in traffic signals. The LCUs use the loop detector signals to determine volume and occupancy and (for dual loop configurations) speed. The LCUs also continually check the loops for long stretches of continuous presence or complete lack of presence, which may indicate loop detector problems. Problems are reported to the mainframe computer at the TOC and can result in reconfiguration of the loops.

## CAPITAL

The Cellular Applied to ITS Tracking and Location (CAPITAL) ITS Operational Test is based on partnerships between the FHWA and several public and private participants. It is a project that was designed to determine if cellular telephone technologies provide an accurate and cost-effective means of wide area traffic detection and surveillance. In addition it will determine if vehicle location information from cellular phone-equipped probe vehicles can be effectively integrated into real-time traffic control system for regional traffic management, traveler information or other purposes. The two basic components of the system are a vehicle locating component and a communication link to pass the information to a traffic control and management center. The CAPITAL system allows any vehicle equipped with a cellular phone to be a potential probe, with infrastructure modifications required only at the cellular tower sites. Unlike the tag-based approach which yields probe data only on those road segments instrumented, CAPITAL can provide probe vehicle speeds every few seconds on any road segment in the cellular coverage area on which mobile phone users are traveling.

The traffic information component is performed by the Traffic Information Center. Within the center, MIST software processes probe vehicle speed data and cellular call activity statistics. Data fusion techniques synthesize real-time estimates of traffic conditions from raw speed data and historical speed profiles. An important finding from the system test is that cellular phone usage on a given link accelerates as traffic volumes approach and exceed link capacity—indications of presence of incidents. A high concentration of Emergency 911 calls was another indication of an incident. The data distribution component provided the capability to broadcast traffic to designated users. Traffic information was displayed graphically at several traffic control centers and other selected locations for in-vehicle use. Land-line and cellular telephones were used to deliver traffic information to users (Robert Larsen, August/September 1996)

## 2.6.6 EVOLVING TECHNOLOGIES

### Smart Call Box

A Field Operation Test that will use California's extensive call box system and increase its functionality by adding an interface to traffic management devices. The project will test the feasibility of using the Smart Call Boxes to collect traffic census data; obtain traffic counts, flows and speeds for incident detection; report information from roadside weather information systems; control changeable message signs; and control roadside closed-circuit television cameras.

### Vehicle Probes

The vehicle itself has become an important surveillance tool for monitoring conditions on the roadway. Vehicles, acting as moving sensors (probes), can provide speed and delay on links traversed. This information can be transmitted to a central computer system where information can be combined or fused with other information for more accurate representation of the actual travel conditions. Technologies that utilize vehicles as probes include:

### Automatic Vehicle Identification (AVI)

These systems permit individual vehicles to be uniquely identified as they pass through a detection area. A roadside communication unit broadcasts an interrogation signal from its antenna. When an AVI equipped vehicle comes within range of the antenna, a transponder (tag) in the vehicle returns that vehicle's identification number to the antenna. This information is then transmitted to a central computer where it is processed.

### Automatic Vehicle Locating (AVL)

This system enables the approximate location of a vehicle to be determined and tracked. There are numerous techniques and technologies that are used for locating a vehicle including Global Positioning Systems (GPS), radio frequency triangulation, proximity beacons, and cellular telephone systems. For the most part, the position of the vehicle is compared to a map database using map matching techniques.

### Cellular Telephone Probe

A vehicle's location can be determined from signals resulting from cellular phone usage within a vehicle, using RF receivers and triangulation techniques. Using this technology in conjunction with map matching algorithms, vehicles can be tracked. This will allow vehicle speeds, as well as travel times to be measured.

### SOURCES:

- *CAPITAL: Using Cellular Phones As Traffic Probes*, August/September 1996, by Robert Larsen, Traffic Technology International.
- *Detection Technology for IVHS, Volume I: Final Report*, December 1996, US Department of Transportation, Federal Highway Administration, Publication No. FHWA-RD-95-100.
- *Field Test of Monitoring of Urban Vehicle Operations Using Non-Intrusive Technologies, Volume 4: Task Two Report, Initial Field Test Results*, May 1996, US Department of Transportation, Federal Highway Administration.
- *Field Test of Monitoring of Urban Vehicle Operations Using Non-Intrusive Technologies, Volume 5: Task Three Report, Extended Field Tests*, December 1996, US Department of Transportation, Federal Highway Administration.
- *Traffic Detector Handbook*, 1990, ITE.

## 2.7 TRAFFIC SURVEILLANCE

Traffic surveillance is the core of any effective advanced traffic-management system. It helps in accumulating accurate and reliable traffic information; identifying and verifying recurring and nonrecurring congestion; identifying severity of problem areas; continual monitoring of traffic over a network; and evaluating the effects of traffic operational improvements. The Freeway Management Handbook discusses the process of planning, designing, and installing traffic-surveillance systems. The following steps make up this process:

- Problem Identification.
- Identification of Partners.
- Establish Goals and Objectives.
- Define Functional Requirements.
- Establish Performance Criteria.
- Define System Architecture.
- Identify and Screen Technologies.
- Plan Development.
- Identify Funding Sources.
- Implementation and Deployment.
- Evaluation.

Section 2.6 Detection, provides further discussion of this process and the steps.

### SOURCES:

- *Urban Freeway Surveillance and Control: The State of the Art*, November 1972.

### 2.7.1 TECHNIQUES/STRATEGIES

#### Closed Circuit Television (CCTV)

CCTV provides a means of visually confirming detected congestion and monitoring freeway operation. This is performed by operators in a central control room who monitor traffic conditions using cameras installed at selected locations on the freeway. CCTV is used in freeway management for various applications including:

- a. *Incident management.* Once an incident has been detected, CCTV is used to verify and confirm the incident, obtain more information as to location and nature (severity) of the incident, and assist in the response to the incident.
- b. *Monitoring of Traffic.* Traffic on the freeway mainline, HOV lanes and their operation, ramp metering operation, operation of changeable message signs and driver response may be monitored.
- c. *Monitoring of Corridors.* CCTV is used to monitor parallel surface streets and signalized intersections along with ramps and metering to verify traffic and capacity for route diversion.
- d. *Safety and Enforcement.* Monitoring for safety and law enforcement is possible at vital traffic areas such as rail-highway grade crossings intersections, and to detect electronic toll collection violators, truck weigh-station violators, etc.

## 2.7.2 TECHNOLOGIES

The following technologies are used for freeway video surveillance:

- CCTV Cameras, Lenses, and Environmental Enclosures.
- CCTV Cameras with Video Image Detectors (VIDS).
- Compressed Video.
- Video Image Processing.

### CCTV Cameras

The modern video camera uses a chip rather than a tube, to pick up the video image. The use of these chips as image pick-up devices has brought about advancements in camera compactness, durability, and performance. This type of camera is described as a Charged Coupled Device chip camera with a solid-state imager.

Types of tube cameras:

- a. *Standard Vision Cameras*, designed to work in even, constantly full light.
- b. *Ultron and Newvicon Cameras*, designed to work in lower light levels and require an auto iris lens.
- c. *Silicon Intensified Target Cameras*, designed to work in extremely low-light conditions

Types of Charged Coupled Device Cameras:

- a. *Complementary Metal Oxide Semiconductor Cameras*, designed to work in well-lit areas. They have high sensitivity to Infrared (IR) light and low sensitivity in low lighting conditions.

## Advantages:

- Good horizontal resolution.
- Low cost.
- High sensitivity to IR provides the camera with the ability to work with IR lighting enhancements.

## Disadvantages:

- High sensitivity of chip to IR makes this a weak choice for use in the bright sun.
- Low sensitivity to low-light conditions makes it a poor choice for nighttime applications. When the available light drops below an acceptable level, a grid-like set of white lines may appear in the center of the picture.

- b. *Interline Transfer Cameras*, designed to work in areas with wider light variations. They are highly sensitive to IR light and have good response in low lighting conditions.

## Advantages:

- Good horizontal resolution.
- With IR enhancement lighting or white-light enhancement lighting, this camera works well in lower light applications.
- High IR lighting response gives the camera the ability to work with IR lighting enhancements.
- Much-improved resistance to vertical transfer smear.

## Disadvantages:

- Even though this camera has better lower light sensitivity than other cameras, it requires enhancement lighting and wider aperture lenses.
- Because of the camera's sensitivity to IR light, an IR cut filter may be required with outdoor applications.

- c. *Frame Transfer Cameras*, like the interline transfer cameras are designed to work in areas with wider light variations. They have good sensitivity to IR light and have a fairly good response in low lighting conditions.

## Advantages:

- Good horizontal resolution.
- Good low-light sensitivity.

- Good sensitivity to IR light gives the camera the ability to work with IR lighting enhancements.

Disadvantages:

- These cameras are considerably more expensive.
- Only a few manufacturers produce this camera, making availability low.
- The camera imager is very sensitive to bright light sources, which can cause bright vertical lines (called vertical transfer smear) in the image.

Chip cameras (charged coupled device) have numerous advantages over the tube cameras for traffic-management applications. Major advantages follow:

- Tube cameras have a problem with burned or retained image when left viewing a static illumination source. When this happens, there is an actual desensitizing of the target area (the photosensitive material where the light is focused) in direct proportion to the light focused on it. As a result, an image of the area that was being monitored remains after the lens is capped or the camera is relocated. When this occurs, the camera must be replaced. The electronic design and solid-state material of chip cameras prevents them from burning or retaining images. For this reason, chip cameras far outlast tube cameras.
- Tube cameras are large, bulky, and emit a lot of heat. The large size of the camera is attributed to the fact that the camera requires large tubes, it requires a lot of circuitry for high voltage to power the tube, and many board-mounted pieces to run the camera. A lot of heat is produced by the camera tube filament and by the large amount of normal circuitry needed to run the camera. Chip cameras, by contrast, are low-voltage units (i.e., the lack of tubes cuts down the amount of electronic circuitry needed and voltage required). Chip cameras are smaller, lighter, and more versatile, and can be installed where ventilation is not accessible, due to their generation of less heat.
- Tube cameras have blooming problems (i.e., exaggeration of a bright area or spot within a video image, such as that caused by a headlight or street lamp). When a tube camera produces an image, a beam of electrons is produced in the tube by an electron gun and is propelled at great rates toward the front target of the tube. As the cathode of the electron gun in the tube wears out, the electron stream used to discharge the target weakens, thus hampering the tube's ability to fully recover the bright areas of a scene, which causes the blooming effect. Chip cameras do not bloom, since they do not have electron guns or target areas.
- The basic design of the tube camera limits the range of imaging enhancements that may be built-in. Tube cameras are capable only of producing a linear videosegment, whereas chip cameras are able to produce a digital video signal directly from the chip. This allows chip cameras to produce such enhancements as built-in zoom, freeze actions, electronic shutter, and digital video motion detection. Peripheral equipment must be used to achieve such enhancements with tube cameras.
- Chip cameras generally produce better image quality than tube cameras. Even though the tube camera has more horizontal lines of resolution than the chip camera, it has a limited gray shade

response and therefore produces images that appear soft to the human eye. Chip cameras have a very high density or contrast output.

The newest camera technology on the market is the Digital Signal Processing (DSP) camera. These cameras are discussed in Chapter 8, Evolving Technologies.

The use of color versus black-and-white cameras is an important consideration. Black-and-white cameras generally provide higher horizontal resolution than color cameras. Resolution is a measurement of the number of television/monitor lines produced by the camera. It is expressed as lines per picture height, or number of lines horizontally. Black-and-white cameras provide 580 horizontal TV lines, while the maximum number of TV lines produced by color cameras is 460. In addition, black-and-white cameras have better low-light performance than color cameras, which makes them better cameras in locations where there is a lack of highway lighting. Black-and-white cameras are most often used in tunnels where there is not much light. In daytime operations, color cameras perform as well as black-and-white cameras.

Color cameras provide important information that is often needed to describe vehicles and to identify license plates for law enforcement. In addition, as technology advances, color cameras are preferred for traffic surveillance, because their images are more pleasing and most people relate to color images.

### **CCTV Camera Lenses**

The CCTV camera lens is used to project the scene on the camera imager for reproduction as a video signal by the camera. A wide variety of lenses of different sizes with different features and characteristics can be used, depending on the application and environment in which the camera is used. The basic parameters affecting the viewing capabilities of a camera are:

- a. *Focal Length.* The focal length of a lens provides the distance from the lens to the imager. As the focal length increases, the image size also increases and magnification occurs. For freeway traffic management, it is necessary to use a zoom lens with wide focal-length range, to provide both a wide angle of view to view all lanes of traffic in both directions at close range and a telephoto view to zoom on an incident up to a half-mile away.
- b. *F-Stop.* F-stop is a measurement of the amount of light that can reach the sensing device through the lens, a function of the focal length divided by the aperture. The higher the F-stop, the less light transmitted through the lens. As a lens is zoomed for viewing distance or the lens aperture is decreased, the F-stop increases. Freeway traffic management applications require that video images be produced in low-light conditions; therefore the lens should have a low F-stop.
- c. *Angle of View.* Angle of view is a measure of the camera's field of view, typically measured in degrees. When a zoom lens is used, the angle of view decreases as the scene width shrinks. Angle of view is a function of the camera imager size, the lens format size, and the focal length range. Once the required angle of view is determined, decision can be made regarding imager size, lens size, and zoom lens focal lengths.

## CCTV Environmental Enclosures

The main purpose of camera enclosures is to protect the camera and from severe weather conditions (e.g., cold, humidity, rain, snow), road moisture, chemicals used on the pavement during snow and ice (e.g., salt), pollution, dirt, dust, etc. The sealed and pressurized camera housing is most often used in traffic-surveillance operations where the camera is mounted high on a pole outside. The Cohu housing, for example, pressurized with 5 psi of dry nitrogen, has desiccant bags and a humidity indicator installed within the enclosure and is made from aluminum tubing. Most camera enclosure units come with a sunshield and waterproof seal to protect the camera and lens assembly from moisture and the heater/defroster system (Appendix I provides a planner for Traffic Management Video System).

## 2.7.3 CASE STUDIES

TransGuide - From the TransGuide ITS Design Report ([transguidewwwserver.datasys.swri.edu:80/articles/designreport.html](http://transguidewwwserver.datasys.swri.edu:80/articles/designreport.html))

### Cameras

The selection of the camera, the lens, and the spacing between the cameras involved several interrelated factors. A very powerful lens system could meet the video requirements with much wider camera spacing than a less powerful lens system. The more expensive cameras provide remote capabilities that allow the operators to make better use of more powerful lens systems. There are tradeoffs between the initial expense of the cameras and the initial expense of the additional communications system equipment required for less expensive cameras, as well as tradeoffs between higher installation and maintenance and cameras with higher initial costs.

Cameras investigated included one-chip low resolution cameras, one-chip high resolution cameras, three-chip high resolution cameras, and digital cameras. The three-chip color cameras selected provide the good color and resolution needed. The highest resolution digital color cameras did not make a noticeable difference in the useable quality of the picture compared to the associated increase in costs. They also lose some effectiveness when digital magnification is used. The lower resolution and less expensive cameras provide noticeably less useful video to the operator.

The camera technology selected was the three-chip, 1/2-inch, high-resolution color Charged Coupled Device (CCD) Frame Interline Transfer (FIT) camera. Most of the controls on the camera are controllable via an RS-232 interface, a standard serial digital interface used to communicate between devices. The user can control the gain and iris, as well as less frequently used characteristics, including various pedestal levels, shutter speeds, black and white balancing, and the detail level of the video camera (DTL). The one major characteristic for which remote control was not specified is the filter. A 5600K filter was manually selected on each camera as it was installed. The specifications required that the camera use at least 772 pixels horizontally and 492 pixels vertically to provide National Television System Committee (NTSC) video with a minimum of 650 lines of horizontal resolution. The specifications required a wide range of gain settings (-3, 0, 3, 6, 9, 15, and 18 dB) to allow the camera to work well in many lighting conditions, including near dark. The FIT three-chip set eliminates many of the streaking effects seen in the Interline Transfer (IT) chips used in some other cameras. The camera selected meets these specifications and is capable of providing high resolution,

color video signals of good quality in most traffic control and emergency response situations. It provides 750 lines of horizontal resolution with a 60 dB signal-to-noise ratio.

### **Lenses**

The selection of the lens system for TransGuide was an integral part of the video equipment selection process. Camera spacing versus lens magnification was one of the primary tradeoffs analyzed in the selection of a lens system.

The camera/lens configurations were required to provide clear video of incidents at the resulting ranges of 1/4 mile, 3/8 mile, and 1/2 mile. The number of cameras required at each spacing was also considered. The results of the investigation led to a choice of 1-mile (nominal) spacing for the cameras. The availability of a lens stack and camera consisting primarily of off-the-shelf components, which could provide useable video at 1/2 mile, was a prime determining factor in the selection of that spacing.

Once the camera spacing was determined, the selection of the lens system necessary to meet the TransGuide requirements at that spacing could be made. The lenses specified for the TransGuide video system are made up of several components, including a standard 2/3-inch lens capable of being remotely zoomed from 9.5 to 15. An adapter, required to allow the 2/3-inch lens to interface with the 1/2-inch camera lens mount, provides an effective magnification of approximately 1.48. A teleconverter with a magnification of 1.5 and an additional extender lens with a magnification of 2.0 are also specified as part of the lens system. The specifications require that the extender be capable of being remotely added to or taken out of the lens configuration. Overall, the specified lens system is capable of providing a magnification of between 21 and 33.3 without the extender and between 42 and 66.6 with the extender in place.

## **2.7.4 EMERGING TECHNOLOGIES**

### **Digital Video Cameras**

From its inception, video has been recorded and transmitted as analog electrical signals. While analog video transmitters and receivers can be built inexpensively, analog video is very expensive to transmit and to store. Further, today's powerful digital computers cannot process analog signals, so analog information cannot be easily sorted, searched or edited.

The transition of video from the analog to the digital domain changes everything. Digital video can be stored and distributed more cheaply than analog video, and digital video can be stored on randomly accessible media such as magnetic disk drives (hard discs) and optical disc media (CDS). Once stored on a randomly accessible media, video becomes an interactive media, allowing video to be used in various applications.

Digital video also dramatically increases transmission efficiency, which means that communications networks, from the public telephone system to coaxial cable television systems to telecommunications satellites, will be able to carry from six to ten times more channels of video programming than was possible before, dramatically increasing consumer choice. The ability to transmit video over the public

switched telephone network will also allow video conferencing, accelerating the work-at-home movement that is changing the way we are all employed ([ccube.dvptechdoc.com/ftrstry/dvd.html](http://ccube.dvptechdoc.com/ftrstry/dvd.html)).

The major benefit of Digital Signal Processing (DSP) cameras is the minimization of analog circuitry which is susceptible to adding noise to the video parameters with temperature. DSP camera technology eliminates discrete components, thus, potentially improving reliability. The component reduction further facilitates smaller, lighter-weight cameras that consume less power. Firmware within the DSP replaces discrete components and facilitates application of advanced video processing algorithms. With processing speeds of DSP increasing, many more advanced signal processing functions will eventually be accomplished in the camera's DSP. Some advantages of DSP cameras include: it stabilizes image from wind gusts and mounting pole vibrations; it facilitates frame-to-frame integration to improve sensitivity; potential improves image quality by reducing analog components which are subject to drift.

## 2.8 DYNAMIC MESSAGE SIGNS (DMS)

Dynamic message signs are traffic-control devices used for traffic warning, regulation, routing, and management. They are designed to affect the behavior of motorists, thus improving the traffic flow, by providing real-time highway-related information. DMSs are the backbone of the traveler information system, improving roadway operations and safety. They display messages informing the motorist of recurring problems, nonrecurring problems, severe environmental and weather conditions, special events, highway priority and HOV lanes, operational characteristics such as toll facilities, exclusive lanes, routing and diverting of traffic, weigh stations, etc.

DMSs consist of three types of real-time signs: advisory, guide, and advance.

*Advisory DMSs* display real-time information on freeway status and advise motorists as to the best course of action. They are used for incident management and special events. These signs are located at freeways, at entrance ramps to the freeway, and on arterial streets approaching the freeway.

*Guide DMSs* may be used to indicate alternate routes during incident management. Although static trailblazer signs are commonly used to divert motorists to alternate routes, DMSs are becoming more and more popular and have been used to display dynamic trailblazers with changeable arrows for guidance.

*Advance DMSs* are used to alert or the motorist to upcoming advisory DMSs. In most cases, this advance sign is static. However, dynamic advance signs have been used to display a simple message alerting the motorist of more information up ahead. This combination of dynamic signs has been very effective in incident management.

### SOURCES:

- *Guidelines on the Use of Changeable Message Signs*, May 1991, Publication No. FHWA-TS-91-002.
- *Guidelines on the Selection and Design of Messages for Changeable Message Signs*, 1992, by Texas Transportation Institute, Research Report 1232-10.

### 2.8.1 TECHNOLOGIES

#### Rotating Drum

Rotating drum signs are the oldest type of changeable message sign. They are the most reliable and most efficient when a limited message is to be displayed. The rotating drum is a mechanical device with a three-sided rotor, to display messages, and uses raised sheet metal or spray-masked letters on a painted aluminum, wood, or translucent plastic background.

Advantages:

- Low power usage - the sign only draws power when changing the state of the rotor.

- Material used is high-intensity scotchlite (same as static signs) and is highly visible in bright conditions.
- Low maintenance - long mean-time to-failure rate.
- Any color may be incorporated into the message.
- Exact shapes of symbols and letter types can be displayed.

### Disadvantages:

- In the process of rotating drum, unwanted messages may be visible for a short period of time.
- Requires external lighting during low visibility periods and at night.
- Limited number of messages that can be displayed (normally four drums with three messages per drum).

### **Reflective Disk Matrix**

Reflective-disk matrix signs consist of standard reflective-disk elements coated with fluorescent yellow tape on the set side and the reset side of the disk element flat back. This type of sign reflects light from some external source such as the sun, headlights, or bottom-mounted lighting. The signs use power only when the disks are rotated or flipped. The reflective-disk elements are electromagnetic and operate on short DC pulses from a microprocessor controller.

### Advantages:

- Low power usage - power is used only when the disks are rotated or flipped.
- In direct sunlight, reflectivity is more visible than illuminated signs.
- The disk elements are large, giving the message contents a large fill area.

### Disadvantages:

- Poor legibility when the sun is behind the sign, due to poor legend contrast.
- Sun and external lighting can cast shadows, covering portions of the legend.
- Non-glare lexan, used in front of the sign, reduces the life of the lexan requiring changing every year to maintain visibility.
- Reflective disks need to be changed every four years due to fading of the fluorescent yellow tape.

## Neon

Neon signs use neon tubing to form legend characters. The display area size, however, limits the number of messages that can be displayed. The neon tubing may be stacked for each message, or each message may be separate on the sign face.

Advantages:

- Enlarged surface displayed area accommodates more messages.
- Two neon tubes per message provide redundancy.
- Red neon pairs is visible under all types of weather conditions.

Disadvantages:

- Large number of messages requires numerous layers of tubing.
- Green neon is less visible for freeway applications.
- Neon signs do not incorporate light dimming and each application must be customized.

## Bulb Matrix (Incandescent)

Bulb matrix signs are made up of incandescent light bulbs used to form characters or graphics. The bulbs are electronically turned on and held on during the period of display.

Advantages:

- Very effective in capturing motorist attention.
- Excellent visibility under all lighting conditions, except in direct sunlight.
- Low capital cost.

Disadvantages:

- Very high power consumption.
- High cost of maintenance - bulbs change about once every year.
- Mean time between failures is high due to high voltage internally and the electronics required to control the sign.
- In direct sunlight, the bulbs have a tendency to washout.

### **Light-Emitting Diode (Clustered)**

Light-Emitting Diode (LED) signs are made up of clusters and are steady-state devices that provide enough intensity when voltage is applied. Available colors are red, green, yellow, and orange.

Advantages:

- Lower power usage - the LEDs require on the order of 20 MA and are multiplexed to lower the overall power.
- High reliability - rated at 100,000 hours of continuous operation (12 years).
- Effective in capturing motorist attention.
- Writing speed is much faster than electromechanical methods.
- No mechanical parts that can wear and fail.

Disadvantages:

The intensity of LEDs is reduced at high temperatures - ventilation is necessary.

Different colors, such as green or amber, may degrade in direct sunlight.

Since LEDs are low voltage, high current is required to power up the sign.

### **Fiber-Optic Matrix (Fixed-Grid)**

Fiber-optic signs use a light radiated from an internal source (halogen lamp) and directed to the sign's viewing face through a bundle of fibers. On the sign face, the points of light (pixels) are arranged in a matrix array. Each point of light that appears on the matrix screen comes from the end of an individual light guide.

Advantages:

- Message changing is almost instantaneous.
- High contrast and luminance levels result in long distance legibility and high clarity.
- Color filters can create almost any color.
- Messages can either be symbols or words.

Disadvantages:

- The message is displayed only when the internal light source is activated.

### **Fiber Optic Matrix With Shutters**

These fiber-optic signs use electro-mechanical shutters and fiber-optic bundles to form changeable pixels. A light source is covered and uncovered by the shutters illuminating the pixels. The viewing face is similar to the bulb matrix sign with the lighted elements being fiber-optic rather than incandescent bulbs.

Advantages:

- Low power use - a light source is not used for every pixel but for multiple pixels (two 50 W halogen lamps for each set of three characters or 105 pixels).
- Light source through the fiber bundles make day and night visibility very good.
- A magnetic memory in each shutter retains the shutterposition indefinitely without control power.

Disadvantages:

- Light sources only have an average life of 8,000 hours, making maintenance high.
- The electronic controls of the shutters, light sources, and backup light sources are complex, making maintenance high.
- The 5x7 matrix array of fibers restricts the presentation of exact symbols and lower case letters.
- The cone of vision produced by the focused fibers restricts viewing.

### **Liquid Crystal Display**

#### **Blankout Signs**

Blankout signs display messages only when the sign is illuminated. The messages are fixed and can be displayed, in entirety or a portion, when illuminated. They are normally neon-type, but can also be of fluorescent lamps behind a cut-out legend, or fiber optics on a fixed grid.

Advantages:

- Exact shapes of symbols may be displayed.
- Message changing is instantaneous.
- Any color combination may be employed.

Disadvantages:

- Very limited in the number of messages and applications.

### **Hybrid-Fiber-Optic/Reflective Disk**

These signs use the standard reflective disks for daytime viewing and fiber-optic pixels for night-time viewing. The fiber-optic pixel is located behind the disk, radiating light through small holes. A fiber-optic light source is used for every three disk modules and is controlled by a photocell and the sign microprocessor controller. The light source is illuminated at all times but visible only when the disk is in the yellow position - "on."

Advantages:

- Disks may be used to display messages in the event of a power failure.
- In direct sunlight, the reflectivity of the disks is better than that of other illuminated types.
- During night-time operation, illumination of the disks is not required since messages are being formed by fiber pixels.
- Utilizes strengths of both fiber-optic and reflective disk technologies.

Disadvantages:

- Reflective disks must be changed every four years due to fading of the fluorescent yellow tape.
- The life of the light source (bulb) is only 8,000 hours.

### **Hybrid - Light Emitting Diode/Reflective Disk**

These signs are the latest in sign technology. Like the fiber-optic/reflective disk technology, the disks are used for daytime viewing and they open up to allow LEDs to be visible to viewers. The LEDs inside the disks are a cluster of up to five high-power amber LEDs. The LEDs are controlled separately from the disks, and either may be activated in the event of the other's failure.

Advantages:

- Disks may to display messages in the event of a power failure.
- In direct sunlight, the reflectivity of the disks is better than that of other illuminated types.
- Utilizes the strengths of both LED and reflective disk technologies.
- Requires power only to change disks and LEDs, which use only 20 MA per element.
- Long life of both disk elements and LEDs.

## **2.8.2 CASE STUDIES**

Most Freeway Management Systems have DMSs; a discussion can be found in Appendix B.

## 2.9 HIGHWAY ADVISORY RADIO (HAR)

The HAR is intended to provide more specific traffic information at key locations more immediately than is possible through traditional commercially broadcast traffic reports. HAR uses either live messages, preselected taped messages, or synthesized messages, based on information from the Traveler Information Database (TID). It is expected that area HAR systems will ultimately employ an automated Voice Response System (VRS), a microcomputer-based system that uses TID information as its basis for operation. The VRS then develops digitized audio messages for dissemination.

### 2.9.1 TECHNIQUES/STRATEGIES

There are two types of HAR systems: Vertical “Whip” Antenna Systems and Induction Cable Antenna Systems.

#### Vertical Antenna Systems

Vertical Antenna Systems use individual antennas or a series of antennas electronically connected together to transmit information. The signal radiates from the antenna in all directions, providing a circular area of transmission. Vertical antenna systems are small, easy to install, and can be placed within several hundred feet of the roadway. They are also less costly to purchase and install than induction cable systems. However, they are subject to damage by weather, accidents, and vandalism. They often require special equipment to ensure that the signal is stable, reliable, and easily tuneable. Also, because the information is broadcast in a circular zone of coverage, the signal may interfere with other coverage zones on the same or adjacent roadways.

#### Induction Cable Antenna Systems

Induction Cable Antenna Systems use a cable installed either under the pavement or adjacent to the roadway. This type of antenna design produces a strong but highly localized signal within a short lateral distance (100 to 150 feet) of the cable. The signal is strong enough to provide full coverage of a multi-lane facility without causing interference to other HAR systems. Also with this system, messages can be individualized by direction of travel. Information is received by the motorists within range of the cable. As a result, interference to other radio systems in the area is minimized. Since the cable must extend the full length of the desired coverage area, induction cable systems are more costly to purchase, install, and maintain. Furthermore, they are not easy to install, especially in built-up areas or on existing facilities, and once installed, they cannot be transported from one location to another.

#### SOURCES:

- Highway Advisory Radio User’s Guide, Publication No. FHWA-RD-80-166.
- Highway Advisory Radio Systems Design Guidelines, Publication No. FHWA-RD-80-167.

## 2.9.2 TECHNOLOGIES

HAR rely upon any of the following technologies for transmission.

### **FCC-Licensed 10-Watt Transmission**

FCC-licensed 10-watt transmission allows a broadcast radius of approximately 3 miles. Any frequency between 530 KHz and 1710 KHz can be used for transmission. However, the FCC has opened up the former dedicated traveler information/HAR frequencies (530 KHz and 1,610 KHz) to commercial broadcasting. Therefore, the potential for interference from commercial broadcasters has increased. Transmission by this means is omnidirectional. Consequently, several adjoining 10-watt transmission zones may encounter interference problems due to radio frequency interference of adjacent transmitters.

### **Low-Power Transmission (Not Requiring FCC Licensing)**

Low-Power Transmission uses 0.1-watt transmitters (not requiring FCC licensing) interconnected and synchronized to form a zone. The zonal configuration allows relatively linear broadcast zones with 1,800-foot spacings between transmitters.

The technology has been tested as part of the Santa Monica Smart Corridor project. An operational test bed is being set up in downtown Los Angeles to evaluate its performance in a high-density urban zone.

Based on the most recent test results, low-power zonal transmission offers less interference and a stronger signal than a more conventional 10-watt transmitter. Transmitters can be adjusted to avoid overlap between zones.

### **Institutional Arrangement With One or More Public Radio Stations**

By this arrangement, a station's regular programming may be preempted for traffic information. One drawback is that it will be broadcasted region-wide, while the information may only involve one or two locations.

### **Leaky Coax**

Leaky Coax technology uses direct buried coaxial cable to provide linear coverage. However, this technology has not yet been utilized and direct burial cable poses a number of maintenance problems, particularly involving construction projects in the vicinity of the cable.

### **FM Side Carrier Allocation (SCA)**

Side Carrier Allocation (SCA) utilizes the "sideband," or space between radio station frequencies, for digital voice broadcasting. Specially equipped radios can tune in the sideband to receive the clear digital signal.

## 2.9.3 CASE STUDIES

Numerous Freeway Management Systems have HAR; a discussion can be found in Appendix B.

## 2.10 COMMUNICATIONS

A freeway management and control system is comprised of many different elements: field components such as signal controllers, ramp meters, field masters, and changeable message signs; surveillance hardware such as detectors and TV cameras; central equipment such as computers, peripherals, and TV monitors; and the human element—operators and maintainers. Each of these components must exchange information with the others. Further, communication system is required to transfer information between the field equipment and the control center. This information may be in the form of video pictures, voice messages, or control and surveillance data (low-speed data).

### SOURCES:

- *Communications Handbook for Traffic Control Systems*, 1993, FHWA Publication No. -93-052, U. S. Department of Transportation, Washington, DC, by Robert L. Gordon, R. A. Reiss, W. M. Dunn, and D. R. Morehead.
- *Internetworking with TCP/IP, Volume I Principles, Protocols, and Architecture*, 3<sup>rd</sup> Edition, 1994, Douglas E. Comer.
- *Multiple Access Communications, Foundations for Emerging Technologies*, 1993, Norman Abramson, Institute of Electrical and Electronics Engineers.
- *National Transportation Communications for ITS Protocol—Overview*, 1995, National Electrical Manufacturers Association.
- *National Transportation Communications for ITS Protocol—Simple Transportation Management Framework*, 1995, National Electrical Manufacturers Association.
- *National Transportation Communications for ITS Protocol—Class B Profile*, 1995, National Electrical Manufacturers Association.
- *Intelligent Vehicle Highway Systems Mobile Communications Guidelines*, September 1993, by Mitre Corp.
- *1993 Wireless Communications Symposium*, technical Papers, 1993, Hewlett Packard.
- *Proceedings of the Second Federal Wireless User's Forum Workshop*, September 1993, US Department of Commerce.
- *IVHS and Vehicle Communications Compendium of Papers Presented at the 1991 SAE Future transportation technology meeting*, Portland Oregon, August 1991, SAE.

## 2.10.1 TECHNIQUES/STRATEGIES

### Transmission

- a. *Data and Voice Transmission*- Data and voice transmission technology is virtually all digital and largely based on the T-Carrier system and the Synchronous Optical Network (SONET) standards in both public and private networks. This generally allows a second manufacturer's equipment to be added to a given site without reconfiguring the hardware.
- b. *Video Transmission* - Full motion video requires 6 MHz of bandwidth per camera, as compared to a bandwidth requirement of only 0.005 MHz for low-speed (1200 baud) data channel. Video is an analog source, while data communications and the T-Carrier and SONET standards are digital. Video transmission is best done over wide-band technologies utilizing dedicated microwave frequencies, coaxial cable, and optical media (fiber optics). Developments include fiber-optic video multiplexers (providing up to 36 video images over a single fiber) and fiber-optic transmission equipment which permit analog video and digital camera control data to be transmitted over the same fiber.

An alternative is to use digital data/voice media along with the means to compress the analog signal into a digital format. A CODer-EnCoder (CODEC) unit can digitize the video image and compress the bandwidth such that real-time images may be transmitted over the telephone lines or other data communications media (e.g., radio, satellite) and back to an analog format for viewing on a television monitor. A reasonable degree of full motion (10 to 30 frames per second) and resolution is maintained, even when there is significant motion in the video image or the camera is being panned or tilted. The best video quality is associated with the highest transmission rates (typically 384 kilobytes per second or higher).

When it is infeasible to install cable or lease cost for fiber is prohibitive, compressed video offers an attractive alternative. Early attempts at transmitting video over voice grade lines (slow scan) were unacceptable for most freeway management applications because the refresh rate was typically 60 seconds (a snapshot). With compressed video techniques, refresh rates of eight to ten frames per second are possible. Although the image appears slightly jerky, it is more than adequate for monitoring freeway operation. The system requires a compression and decompression computer for each camera/monitor link and off-the-shelf software. The communications medium is a leased Integrated Service Digital Network line, typically \$50 to \$75 per month. Lease cost and installation charges to continue to decrease due to strong consumer demand from PC users.

### Architecture

The architecture of a communications network defines what kind of equipment will be used in the system. It does not necessarily define the medium; the same architecture can often be operated over different media, and different media are frequently used in the same system. In some cases, however, the selection of a particular architecture requires a specific medium or, at the very least, allows or precludes classes of media.

In general, architectural approaches to communications systems fall into two camps. The first camp accommodates systems that require reliable real-time transmission of messages. The second camp does not. The need to accommodate real-time communications therefore becomes the single most important decision in the selection of a communications architecture.

a. *Real Time*

The term *real time* is often defined as messages being passed within a specified amount of time. To most traffic engineers, this time frame is usually one second. Other branches of the computer industry, however, have very different standards. In industrial control systems, for example, real-time operation often requires response to a stimulus within 100 microseconds. Some process-control systems boast of response times as short as 6 microseconds. The use of the term *real time* to describe a desired response horizon is therefore not very specific.

A much more rigorous definition of real time is that the system must respond to external stimuli in a definable and predictable way; that is, deterministically. In a traffic system, for example, the very long one-second event horizon is satisfactory, but the system must predictably provide that response horizon every time. In any system that processes stochastic messages, this predictability is very hard to guarantee. The Internet provides a good example. Internet communications are entirely stochastic. The demand on the network for passing messages depends entirely on the behavior of a host of independent users. So demand on the network is a random variable that can be characterized by a bell curve of some sort. No matter how much communications capacity is provided, there will be some occasions where it is not enough and the required horizon is violated.

b. *Deterministic Architecture*

In systems requiring mandatory real-time communications, a different approach is usually adopted to ensure real-time performance. In these systems, the communications paths are divided, either physically or logically, into private channels with predetermined performance characteristics.

If, for example, a proposed system is required to send a command message to each of 1,000 traffic signals exactly once each second, the private-channel approach will be used. In such a system, traffic signals are grouped into communications subgroups. Each subgroup has its own physical path back to a hub site. Communications on that physical path are subdivided in a very deterministic way. If there are eight traffic signals on that path, the time available on that path will be divided into eight time slices. Each time slice is a predetermined size, and each traffic signal receives exactly that size portion of the physical path no matter what. Thus, each device has a fixed and predictable communications rate. For example, typical traffic signal systems employing private-channel architectures allow eight controllers sharing a single physical channel that has a capacity of 1,200 bits per second (bps). The effective communications rate to each device is therefore 150 bps.

The physical paths are brought back to a hub, and combined onto a trunk line. The trunk line is divided in exactly the same way as the physical path to the traffic signals. On the trunk line, therefore, we may have 64 or 128 traffic signals communicating, each on its predetermined time slice. The overall performance of the trunk line has to be fast enough to accommodate all the fixed private

channels. This process of devoting time-slices to each of a large number of private channels is called time-division multiplexing.

The private-channel scheme was developed by the telephone industry. When a caller places a call, a private channel is established between the caller and another party. That channel may traverse a huge variety of distribution lines, trunk lines, and switching equipment, but the size and capacity of that channel is predetermined. Virtually all technologies developed within the telephone industry use time-division multiplexing, including the T-Carrier standards for twisted-pair wire and the SONET standards for fiber-optic cable.

The advantage of a private-channel architecture is that it is deterministic, and can therefore accommodate systems that require real-time communications. The specific time horizon of the real-time messages is not important, architecturally. Once a deterministic system is chosen, then the time horizon is strictly a matter of the performance of the network. Faster communications channels can move more messages, and can therefore either be divided into more private channels or can deliver a higher message frequency to fewer channels.

The disadvantage of the private-channel scheme is that it does not efficiently use communications capacity. Each remote device must deterministically be allotted its own time slice, whether or not a message is being communicated. During the many intervals when a device is communicating nothing useful, the sub-channel devoted to that device is wasted. The corollary is that a device that is transmitting many communications must meter those messages out over many time slices. Private-channel systems cannot flexibly accommodate occasional high demand from individual devices.

Deterministic architecture using private channels is analogous to the circuit switching architectures described in many communications textbooks.

### c. *Stochastic Architecture*

For systems that do not require real-time communications, a different approach may be taken. The shared-channel approach allows many devices to share the same communications channel. In the previous example, eight intersections each received a private channel of 150 bps. In a shared-channel system, all eight intersections use a single channel of 1,200 bps. Messages are routed to the proper controller by including its address in the message. This approach is known as a multi-drop communications channel, because the physical channel has multiple drops to individual field devices.

Many times, the shared channels will be returned to hubs and be multiplexed onto private-channel trunk lines to make a hybrid system. In other cases, the hub computers combine the data from many devices, and then share a high-capacity channel back to the central computer for a multi-tiered shared-channel architecture. By the shared-channel approach, messages that have the appropriate addressing can traverse quite complicated networks to get from origin to destination. This is the approach taken in most computer networks and inter-networks, including the Internet.

The advantage of the shared-channel approach is that it can be very efficient. Because each controller has access to the full capacity of the channel, each controller can send high-demand bursts of communications at high speed, though, of course, doing so will inhibit communications to other

devices on the same channel. Demand on the channel increases when more than one device needs to send large messages at the same time. The time required to get the messages through becomes longer in this case. The channel capacity must be designed, therefore, to provide acceptable time horizons under expected peak communications demand.

Another advantage is that once the decision has been made to avoid mandatory real-time communications, media with reduced reliability become available. For example, a system that requires real-time communications for sending control commands may not be able to tolerate the indeterminacy of a shared-channel communications scheme. Such indeterminacy may also be caused by limited communications reliability. Many wireless communications media provide no better than 99.9 or 99.99 percent reliability, which translates to many minutes of downtime a month. This reliability has nothing to do with the equipment; it only describes loss of communications caused by external events, such as bad weather or outside radio frequency interference. Systems that cannot tolerate indeterminacy will need determinacy in the media as well as determinacy in the architecture.

A disadvantage of shared-channel schemes is that they are more difficult to design. In determinant systems, the communications demand generated by each device is fixed. The capacity required to accommodate that fixed demand is provided to every field device at all levels of the system. The system therefore behaves in a way that can be determined beforehand. In stochastic systems, communications demand is a random variable that must be characterized before the appropriate capacity can be designed and provided. The reliability of the system is a function of the probability that the capacity will be great enough for any given demand peak. Fortunately, most traffic engineers should implicitly understand this approach: it is exactly the same way roadway capacity is characterized.

Because messages in nondeterministic systems must move in a shared environment, each message must sufficiently describe itself to the communication system to get to the other end without mishap. The protocols required to accomplish this are discussed later, but this requirement suggests that all data in shared-channel architectures be divided into messages, or packets. These systems are therefore often described in communications textbooks as packet-switched networks. Not all shared-channel architectures use packet switching in the rigorous sense, but the concept is an important one. A designer seeking very detailed technical information on how to design stochastic communications systems will find the needed information within the packet-switching subject area.

#### d. *The Line is Not So Clear*

Having established a sharp line between the two architectural approaches to communications systems, that line must be rendered again with much less clarity. Both camps (those that accommodate real-time transmission and those that don't) have been migrating towards one another for some years, and elements of each will be found in most modern systems.

In stochastic systems, we assume that the messages are generated by a large number of independent sources, which makes the total communications demand a random variable that is close to being normally distributed. This demand pattern creates a very symmetrical bell curve centered on average demand. In most traffic-management systems, the demand distribution is not the result of a large number of independent processes. In traffic-signal systems, for example, the local controllers send and receive very similar messages defined from a very limited set of possibilities. Peak demand is

much easier to control and predict, such that it is not completely indeterminate. If a system can impose sufficient discipline on the peakiness of the communication demand, the peak capacity can be characterized and designed for. Consequently, even a stochastic system can provide determinant real-time communications when carefully designed. This possibility allows shared-channel communications architectures to be coupled with centralized control systems normally used with private-channel determinant communications systems. For example, the Siemens UTC, which is the parent system of one of the implementations of the highly centralized SCOOT system, employs shared-channel multi-dropped communications between the communications processors and the local intersection controllers.

Likewise, many systems that do not require real-time communications use determinant communications architectures. Traditionally, the telephone industry has driven advances in communications technology. Consequently, equipment for private-channel communications systems is readily available, and has been for many years. The use of such an approach is a practical response to the marketplace, even though it may not be architecturally pure. With recent advances in communications for computer networks, and with much greater interest from the telephone industry in stochastic networks (at least for major trunking), practical reliance on private-channel equipment and standards is no longer necessary.

Finally, it is quite possible and common to implement one system architecture via a medium designed for the other architecture. For example, many fiber-optics systems use SONET, which has a private-channel architecture, to move shared-channel communications. Many computer networks use a T-carrier standard to carry a stochastic communications protocol. Again, this is a response to the availability of equipment designed by and for the telephone industry. This mixture of approaches is also associated with the tendency of stochastic systems to depend on software standards and deterministic systems to be rooted in hardware standards. In stochastic shared-channel architectures, discipline is maintained by a software communications protocol, such as the TCP/IP suite used by the Internet. These protocols may be communicated over telephone lines, which are the ultimate in private-channel communications media. But this is an architectural anomaly. Any stochastic system using TCP/IP would much rather use a physical medium that is oriented towards shared channels, such as Ethernet 10-Base-T, which also runs on twisted-pair wire (at least for a limited distance).

A detailed example will summarize these principles. A single physical communication path on a trunk line may carry 10 T1 channels. Each T1 channel can carry 24 telephone conversations at one time. So the physical trunk line that carries 10 T1s can accommodate 240 simultaneous conversations. If each of these conversations is actually an Internet user dialed into the system, then the performance of each of these users will be a guaranteed 56,000 bps, using the latest modems. This private-channel scheme provides a deterministic 56,000 bps to each user. A single Ethernet network can provide computer network access for 240 users. When all the users are demanding the network, the performance might be quite poor, much less than the 56,000 bps throughput that the private-channel scheme would guarantee. Even though the overall capacity is about the same, the shared Ethernet channel would be more clogged because of the extra time required to sort out collisions and deal with the chaos of randomly generated messages. Most of the time, however, only a few users would be generating high demand, and those users would get much greater throughput than 56,000 bps. A single user accessing the trunk line over the dial-up connection is limited to 56,000

bps, even if the other 239 channels are not being used. A single user on the Ethernet network gets 10 million bps.

e. *Reliability and Redundancy*

Many communications systems enhance reliability by providing redundant elements. The most common redundancy scheme for communications is the self-healing ring, which is defined for fiber-optics systems that employ the SONET standards. In a self-healing ring, signals are transmitted simultaneously clockwise and counterclockwise around a ring of hubs. Each hub is therefore connected to the other hubs in two ways. If the cable is broken between two hubs, then all the hubs are still accessible from one or the other direction of the ring. The equipment in a SONET system is designed to automatically switch between directions to correct faults.

Some systems design self-healing rings along a linear path by squeezing the ring together until it forms a line. This does not provide protection as well as a true ring, because a physical interruption of the cable will likely affect both sides of the "ring" and still isolate some of the hubs. However, it still allows the system to recover from equipment failures. This approach is used in the San Antonio TransGuide system, adding protection as the two sides of the ring travel down opposite frontage roads of the freeways that are part of the system.

Ring topology is commonly associated with SONETs, but the concept may be applied anywhere. For example, the new Las Vegas Area Computer Traffic System uses a broadband microwave communications network. Half the band is transmitted clockwise, and the other half is transmitted counterclockwise. When interference disturbs a link, all hubs are still accessible, though the time to switch from one side to the other may be longer than would be typical in a SONET.

Redundancy helps improve reliability, but it does not reduce maintenance requirements. Actually, ring topology and other redundancy schemes increases maintenance requirements, because the equipment is more complicated, and in some cases, duplicated (which means that more equipment must be maintained). Redundancy increases overall maintenance requirements, but reduces the urgency of maintenance. When a link is damaged on a SONET, the system continues to function, so operation of the system does not depend on immediate repairs. The repairs must still be made quickly, because the system no longer has redundancy. In many agencies, maintenance budgets are down to crisis levels, meaning that only those repairs that are critical to system functionality will be performed. In such circumstances, ring redundancy may be a problem rather than a solution, because the redundant elements will not be repaired as they fail. Once the redundant elements have failed, the system no longer benefits from the enhanced reliability. But the agency has still borne the increased cost of redundancy, both in initial cost and in overall maintenance requirements. In such cases, agencies would be better served spending the money on systems that are easier to maintain and resist damage more effectively.

An example illustrates the point. Some systems being designed achieve redundancy by connecting local devices to two communications hubs using fiber optic cable. If a local device is destroyed (as the result, for example, of a traffic accident) then the other devices on the path will still be functional. Thus, an agency might use a fiber system, knowing that the agency's inability to perform quick repairs on the fiber is offset by the redundancy scheme's reduction of the urgency of repair. If that

agency suffers from crisis-level maintenance cutbacks, however, then the minimum repair necessary to make the system functional may be all that is performed. Continuing the example, if a backhoe dips into the cable between two devices, the repair may be deferred indefinitely, because the system is still operational. In such a case, the agency would have been better off with a simple twisted-pair copper system that can be more easily repaired with typically available technicians. The savings can be directed towards deeper trenches and larger conduit, which will make damage less likely and repairs easier.

Another principle that emerges from this example is that the technology of the communications should not outstrip the ability of the maintaining agency. Returning to the example, a technician performing the emergency replacement of a knocked-down traffic controller will attempt a repair of the twisted-pair copper wire. The wire itself will also be more robust and resistant to damage. If the medium is fiber-optic cable, then the fiber may be damaged some distance away from the place where it comes out of the ground, and a broken fiber at the connector will require a specialized repair. Consequently, the technician performing the emergency controller replacement will most likely not even attempt a repair, and will defer it to a more specialized repair crew at a later time. System reliability therefore suffers.

**Guidelines and Standards**

a. *Deterministic Architecture*

Private-channel architectures were initially developed by the telephone industry. Consequently, a host of standards are available for defining multiplexed digital communications. All these standards can be characterized in a single continuum known as the electrical digital hierarchy, which is represented in Table 2.10.1 below.

Table 2.10.1. The Electrical Digital Hierarchy.

DESCRIPTION	TOTAL DATA TRANSFER RATE	COMPOSITION
OC48	2.48832 Gb/s	48 DS3 (1344 DS0)
OC36	1.86624 Gb/s	36 DS3
OC24	1.24406 Gb/s	24 DS3
OC18	933.12 Mb/s	18 DS3
OC12	622.08 Mb/s	12 DS3
OC9	466.56 Mb/s	9 DS3
OC3	155.53 Mb/s	3 DS3
OC1	51.84 Mb/s	1 DS3 (plus SONET overhead)
DS3	44.736 Mb/s	28 DS1
DS2	6.312 Mb/s	4 DS1
DS1C	3.152 Mb/s	2 DS1
DS1(T1)	1.544 Mb/s	24 DS0
DS0	64 kb/s	Basic Unit—One Digital Voice Channel
DS0A	64 kb/s	2.4, 4.8, 9.6, 19.2, or 56 kb/s filling a single DS0
DS0B	64 kb/s	20-2.4, 10-4.8, or 5-9.6 kb/s subchannels on a single DS0

The hierarchy is divided generally into three sections. One of these standards covering standard serial communications, sets guidelines for communications at the slowest speeds. These standards are defined by several agencies, including the Electronics Industries Association (for physical media) and the Consultative Committee for International Telephone and Telegraph (CCITT, for communications layers above the physical media). Communications at this level range in speed from 110 bps to 115,000 bps, but multiplexing equipment is only available to combine the speeds shown for DS0A and DS0B.

Any discussion of serial communications standards also requires a discussion of modem standards. Most modems designed for use on dial-up telephone lines have a standard serial interface. On the telephone line side of the modem, the standards for how the digital signal is modulated into an audio signal vary widely. At the slowest (and oldest) end of the spectrum, the standards are defined by the telephone companies, and designate how signals will be modulated so that two modems from different manufacturers can interoperate. These standards include the Bell 103 standard for 300-bps communications, and the Bell 212 standard, which increased the speed to 1,200 bps. The later standards include V.22bis (for 2,400 bps), V.32 (for 9,600 bps), V.32bis (for 14,400 bps, or slower if need be), and V.34 (for 28,800 bps and faster, with fallback to slower speeds when necessary). In addition to modulation standards, other standards are usually applied to improve robustness and throughput. These include V.42, which is for error correction, including the Microcom Networking Protocol (MNP) classes 1 through 4, and V.42bis (MNP 5), which is a data compression standard used in conjunction with V.42. All the "V" standards are defined by the CCITT.

The middle level of the electrical digital hierarchy covers high-capacity communications over copper-wire links. These are commonly known as the T1 or T-Carrier standards, but are more specifically known as the Digital Signal Types. "T1" is also used to describe DS1, or Digital Signal Type 1. The T1 standards grew out of the telephone industry, and are currently defined by the American National Standards Institute (ANSI). The basic unit of T1 is a digital voice channel, which is defined nominally to be 64 kilobits/second, and is known as DS0 sixty-four kilobits/second is the digital capacity required to carry the human voice with reasonable fidelity for telephone applications. A DS0 can be further subdivided, but any signal larger than DS0, all the way up through SONET speeds, must first be broken down into DS0 channels, sent across as many channels are required, and reassembled on the other end. Higher standards are multiplexed combinations of multiple DS0 channels.

If the basic unit of transmission is a DS0 voice channel, then the basic unit of trunking, or multiplexing, is the DS1, which is commonly known as T1. A DS1 multiplexes 24 voice channels. The highest level in the T1 standard is DS3, which combines 28 DS1s, or 672 voice channels.

A designer using T1 has much equipment to choose from. For example, a number of manufacturers provide Ethernet bridging equipment that runs over T1. The full Ethernet speed of 10 Mbits/second cannot be obtained, but the equipment handles the task of dividing the Ethernet signals into 24 DS0s, and then recombining them on the other end. Similar equipment is available for moving video images across multiple voice channels. So, even though T1 is a private-channel standard, shared-channel communications schemes will often employ it at some level.

A T1 is also available as an unchannelized path. The Ethernet bridging equipment mentioned in the previous paragraph uses unchannelized T1, which means that the equipment is designed to use the full bandwidth of the T1 as if it was a combination of many DS0s.

At the highest level, the electrical digital hierarchy defines standards for fiber-optic systems. The ANSI T1 standard body set guidelines for SONET to carry multiple T1 combinations across linked self-healing rings. Outside North America, the same standards are called Synchronous Digital Hierarchy by the CCITT. The basic unit of SONET is the optical carrier. The smallest unit is the OC-1, which carries (in addition to the SONET overhead) a DS-3. SONETs are defined up to OC-192, but equipment is commonly available only for OC-3, OC-12, OC-24, and OC-48. OC-48 carries 48 DS-3's, which is equal to 32,256 voice channels, on a single pair of fibers.

SONET includes a definition of self-healing rings, as discussed in the previous section.

### b. *Nondeterministic Architectures*

Standards for nondeterministic, or stochastic, architectures must include the control protocols necessary to resolve conflicts within the system. The protocols must also have the capability to route messages through the network, because messages for different destinations will share the same communications path.

For systems that can impose a lot of discipline on the variety of messages, the protocol can be quite simple. The greater the need for variety, the greater demand on the protocol to manage the transmission of messages.

The need to manage the transmission of messages through a shared-channel network means that communications standards must include software as well as hardware requirements. The software requirements will be discussed in the next section on protocols.

The hardware standards for shared-channel communications are mostly generated by the computer networking industry. For low-speed communications, these standards are the same as those used in private-channel architectures. For example, many traffic-signal systems use 1,200-bps modems conforming to the Bell 212 modulation standard. This standard allows these modems to be multi-dropped along a common pair of copper-twisted pairs.

For high-speed twisted-pair networks, Ethernet is the protocol of choice. Ethernet defines a basic communication rate of 10 Mbits/second, and requires synchronization on both ends. This data rate and synchronization limits the time delay for a message to 44 microseconds. The physical range for Ethernet is therefore sharply limited, and the standard is therefore used mostly for local-area networks. Ethernet is a multi-dropped environment, where all devices hear all messages. Ethernet has recently been extended to 100 Mbits/second.

When network segments that use the same software protocols need to be joined together over a long distance, a *bridge* is employed. A bridge packages data from one subnetwork and sends it over a fixed path, such as a telephone line, to a corresponding bridge attached to another network. The bridge makes all the members of one network look like local members of the other network, and vice versa.

When several networks that use different software protocols are joined together, a *router* manages the mapping of messages from one network to the other. The router looks more deeply into the message than does the bridge, and repackages the message to send it on to the destination or the next router. When routers are used to build links over fixed paths, they are called *bridge-routers*, or *brouters*.

The Ethernet standard (IEEE 802.3) is controlled by the Institute of Electrical and Electronics Engineers, but the standard was originally defined by a consortium of Intel, DEC, and Xerox.

For networks of high-performance computers that are linked with fiber-optic cable, the Fiber Distribution Data Interface (FDDI) is becoming an accepted standard. FDDI provides a token-ring approach protocol running at 100 Mbits/second. In a token ring environment, a token is passed from computer to computer in the network. While the computer holds the token, it may transmit packets of data to other computers. The token is passed in turn around the network repeatedly. The token-ring approach is deterministic in that discipline on the channel is absolute, and all possible collisions are precluded. FDDI is defined only for optical networks, and specifies the 1,300-nm wavelength of light. FDDI networks can be repeated to cover up to 200 km.

### c. *Polled and Non-Polled Protocols*

Multi-dropping creates the possibility of two devices attempting to use the channel at the same time. This potential conflict requires some discipline in how devices can access the channel. The simplest scheme precludes remote devices from transmitting a message unless they are individually commanded to do so by a master device. This approach is known as a *polled protocol*. The master device sends out a poll request message to a specific field device. That device then sends a response. Other devices will hear the transaction (or at least the master's side of the transaction), but will keep silent. The polled protocol can be characterized by the phrase *speak only when spoken to*.

When modems prepare a communications channel for communications, they first send out a tone called a *carrier*. The modem will then vary the pitch or frequency of that tone to distinguish between binary ones and zeros (called *marks* and *spaces*). Once the message is complete, the modem will stop sending the carrier tone, leaving the channel silent.

1. CSMA. Non-polled protocols are also known as multiple-access protocols, because they permit various devices on the channel to initiate communications protocol. The simplest of the non-polled protocols is known as Carrier-Sensed Multiple Access, or CSMA. With CSMA, a modem will first listen for a carrier tone on the channel. Hearing none, it will establish a carrier. The problem comes when two devices try to establish a carrier at the same time. The messages from both devices will be on top of each other, and both will be garbled. The receiver will not receive the message, and will send no acknowledgment back to the sender. In most CSMA protocols, the modems will wait a short period of time and, hearing no acknowledgment, attempt the communication again. The CSMA approach can be described by the phrase *speak when the channel is clear*.

The problem with CSMA is that when a large number of devices is trying to access the channel at the same time, the performance of the channel can spiral out of control. With a constant delay before resending a message, two devices can compete for the channel repeatedly

until one gets ahead of the other. The more crowded the channel, the more likely these collisions occur, and the greater the backlog of sent-but-unacknowledged messages, which means that the devices will try harder to access the channel, and so on until the channel is hopelessly jammed. CSMA systems, therefore, work best in situations where some external discipline might be imposed on the communications. For example, many traffic-signal systems temporarily remove an intersection from the system if three successive status messages are missed. In a non-pollled environment, such a rule will allow the system to back out of jammed state, although only by reducing the operational performance of the system. These problems are avoided by placing strict limits on the loading of the channel. These limits usually restrict the average loading to about 10 percent of the channel, assuming independent message generation by all users. When message generation can be more carefully controlled, greater loading is possible.

Some of the schemes for recovering from collision are very clever and effective. Well-disciplined CSMA systems employing these schemes can achieve channel loading as high as 85 percent without jamming.

CSMA is most commonly used in high-performance networks where the ratio of individual user need is small compared to the total capacity. For example, Ethernet uses CSMA. Ethernet subject to jamming (called Ethernet meltdown) and depends on the software protocols above the Ethernet to resolve high-conflict rates.

2. *ALOHA*. To minimize the problem of jamming for networks where users access the channel at random times, devices can wait a random period of time before attempting to resend unacknowledged message packets. The random delay injects a little air into the jam, and allows it to sort itself out much more efficiently. These random delays also decrease the efficiency of the channel to levels much lower than well-disciplined CSMA systems. ALOHA systems were developed primarily for satellite communications, but are also employed in many data radio systems.
3. *Deterministic Multiple Access Systems*. Some systems impose absolute discipline on the channel to prevent any possible collision. As previously mentioned, the token-ring approach is one such system. Others include Frequency Division Multiple Access (FDMA), which gives each user its own frequency, and Coded Division Multiple Access, in which the signaling characteristics of each packet are uniquely coded such that other packets become invisible to the equipment. Spread-spectrum radio uses CDMA.

d. *Software Protocols*

Most of the previous discussion centers on the specific protocols of the communications medium. All these standards are applied at the bottom layer of a communications system. The bottom layer is the physical layer, the part that directly manages the signaling on the medium in question. Once the signals are translated into bits and bytes, they are processed by computer programs within the communications software of the device. Each message is encapsulated within a series of wrappers, each of which defines some aspect of the packet of data. These wrappers are called layers, and conform to a communications model developed by the International Standards Organization (ISO).

The ISO seven-layer model has been called the Open Systems Interconnect (OSI) model. The ISO-OSI model is shown below:

Layer	Description
Application	The program that desires the communication resides here. For example, the Internet uses the Simple Mail Transfer Protocol to send e-mail and the File Transfer Protocol. Another example is the Internet's Hypertext Transfer Protocol, used on the World Wide Web.
Presentation	Used for providing services needed by many applications that may share a network, such as conversion of graphical images into bit-streams or text compression. Not used much in actual systems—most of the functionality originally modeled for this layer has ended up in the application layer.
Session	This layer was specifically designed to handle remote terminal access services. Not used much in public systems.
Transport	The Transport Layer ensures end-to-end reliability. Protocols within this layer break outgoing messages into packets, and add ordering and error-correction information to each packet. For incoming messages, the transport protocol reassembles the packets into continuous data streams. The transport layer makes sure that packets are received intact and assembled in the proper order, which is called guaranteed delivery and sequential ordering, respectively.
Network	This layer is the boundary between logical addresses and physical addresses. For incoming messages, this protocol determines if the message belongs on this computer. For outgoing messages, this protocol wraps a routing header onto the message to send it to the next routing device or to the destination, if it is directly connected. The Internet Protocol is a network layer protocol.
Data Link	The interface between the upper software layers and the hardware. The data link protocol defines the boundaries of message packets to the physical hardware, and includes a simple acknowledgment and error detection scheme to confirm successful communications to the hardware.
Physical	The physical layer defines the signaling and timing characteristics of the medium over which messages will be transmitted. All modem protocols, for example, and all the protocols of the Electrical Digital Hierarchy reside within the physical layer.

A central principle of the multi-layer model is that each layer need not know anything about the other layers. The application layer does not need to know anything about which protocol is used in the transport layer. Consequently, a profile of protocols can be defined, selecting the specific protocol for each layer as appropriate for the system in question.

An example shows how this system works. In the Internet, a user of the World Wide Web will use a specific profile for downloading a Web page. At the application layer, the protocol will be the Hypertext Transfer Protocol (HTTP). The presentation and session layers are not used by the Internet. The transport layer protocol will be the Transmission Communications Protocol, or TCP, which converts outgoing data streams into packets, or datagrams, and adds to each packet the information needed to guarantee delivery and reassemble the packets in the proper order, which the TCP module on the other end will do. The TCP module cares not that the application protocol is HTTP. Thus, the TCP module in the origin computer talks to the TCP module in the destination computer.

Carrying the example further, the TCP datagram is then passed to the network layer, in which resides the Internet Protocol (IP). The IP will add routing information, including the destination logical address. The IP will then consult its routing tables, and determine which computer the datagram needs to go to next, and will add that address to the datagram. Again, the IP knows nothing about the information added or needed by the TCP or the HTTP. The IP module transmits information to the IP module *on the next machine in the network*. Thus, end-to-end matters are handled by the TCP module, and machine-to-machine matters are handled by the IP module.

Finally, the IP-massaged datagram is passed to the data link layer for interfacing with the hardware. For dial-up users on the Internet, this layer is probably controlled by the Point-to-Point Protocol (PPP). The PPP establishes a channel over which two modems may communicate, and ensures error-free transmission to the modem. The PPP module in one computer talks to the PPP module in the next computer on the network, and is again ignorant of all protocol packaging above it.

The suite of protocols used by the Internet is known as TCP/IP, but includes many more specific protocols than just the two most popular ones listed in the name.

This discussion of multi-layer protocols is critical to understanding the National Transportation Communications for ITS Protocol (NTCIP) discussed in a later section.

### e. Video Communications

The data transmission requirements for full-motion video are so extensive that they deserve a separate discussion. The addition of video surveillance to a system design fundamentally changes the requirements of the system. For data only, the major decision concerns selecting deterministic digital communications or nondeterministic (stochastic) digital communications. When video is added, the major decision affecting cost is whether or not to use analog video transmission or digital video transmission.

1. *Analog Video*. Full-motion video, in analog form, requires 6 MHz of bandwidth, of which 4.5 MHz is required by the color-burst, or picture, and the remainder for audio. Analog video is a continuous scan of electrons moving horizontally in rows down the screen. In North America, video signaling is defined by the National Television Standards Committee (NTSC). NTSC video is defined as 525 lines of vertical resolution, or scan rows, and all rows are scanned 30 times per second. In many systems, including most consumer television systems, the signal strength and display equipment is not good enough to clearly transmit and display all 525 lines, and most consumers are more accustomed to seeing about half that. The standard is defined to allow this degradation without loss of synchronization.

Television stations broadcast television using amplitude modulation, or AM. Television can also be transmitted using frequency modulation, or FM. Although broadcast television uses only AM, video surveillance systems have other options.

In private microwave systems, video can be broadcast either by AM or FM signals. Two bands (12 GHz and 18 GHz) are reserved for AM wideband transmission. *Wideband* refers to the transmis-

sion of many video images, each of which has its own frequency, using a single transmitter. Many other frequencies are available for FM video transmission.

FM television may be transmitted over fiber optics very effectively. In fact, in most systems this proves to be a much cheaper alternative than digital video. Multiple FM video signals can be modulated onto a single fiber, routinely allowing as many as 32 video channels per fiber. Specialized equipment can transmit as many as 100 video channels using FM multiplexing.

Many systems are now using equipment that transmits analog video over twisted-pair copper wire for distances of up to several miles, with repeaters. Video is also commonly transmitted over coaxial cable, though this application seems less popular than it once was.

In any medium, however, the signals themselves are governed by the NTSC description.

The advantages of analog video transmission are it provides extremely high-quality video cheaply and efficiently over wireless and fixed media. For example, a fiber system moving video images needs a modulator for each image, an FM multiplexor, the fiber, a de-multiplexor, and a demodulator between the camera and the control-room display switcher. Each of these items is routinely available, and is commonly used for video surveillance in other industries and for transmission by cable television companies.

Another advantage is that it allows graceful failure. If the communications link is degraded, the picture will be degraded in like amount. A significant amount of reserve signal quality is available in analog video. For example, video at the camera meets commercial broadcast standards, and has a signal-to-base noise ratio of about 58-60 dB. The decibal scale is logarithmic: each change of 3 dB signifies a doubling or halving of signal strength). Most cable television systems operate at about 40 dB signal/noise, which is about 1.5 percent of the signal strength of true broadcast quality. Yet most viewers would consider the resulting image to be good. Usable video can be obtained with signal-to-noise ratios of as little as 28 dB.

The disadvantage of analog video is that it does not usually permit sharing a channel with other digital communications. Generally, a fiber that is devoted to analog video cannot be used for digital communications, and vice-versa. This limitation may not be total, however. Equipment is now available for using different wavelengths of light on the fiber. One wavelength is reserved for data, and the other for video. This technique is known as wave-division multiplexing.

2. *Digital Video.* If each horizontal scan line is divided into 700 picture elements, or *pixels*, then each scanned screen image contains 367,000 pixels. If each pixel defines a level of red, green, and blue (the additive primary colors used in constructing a color video image), and if each color can be defined at 256 levels, then each pixel needs three bytes to describe it. The size, in bytes, of a single frame of broadcast-quality video is therefore a little more than a megabyte. If 30 frames are transmitted per second, then the data stream is going by at 265 Mb/second. In practice, an analog video image can be stream-digitized at less than broadcast quality at much lower rates. Typically, full-motion video is expected to consume 100 Mb/second. Even at this rate, the bandwidth requirements of digital video are vast compared to data requirements for most systems.

The computer industry, driven by developments in video teleconferencing over telephone lines, has been working on strategies to reduce these requirements to more acceptable levels. These strategies fall into three categories: image compression, data compression, and reduced framing rates.

Image compression reduces the resolution of the image, both in terms of sharpness and color. For example, a digital video image that is 263 lines by 350 pixels and monochrome can be transmitted in about 23 Mb/second. This is the same quality as most black-and-white television. If the monochrome image size is further reduced to 128 by 170, as with many teleconferencing systems, the requirements are further reduced to about 5 Mb/second. Full-range monochrome can be reasonably represented with 256 shades of gray, or a single byte per pixel. We can also transmit color images with a single byte per pixel, but the quality of the color will be noticeably reduced. Still, for many applications, the degraded signal serves the purpose. This is about the minimum image quality that is usable in traffic surveillance systems.

The second strategy is to compress the data as a stream. Most video images contain large areas of single colors that can be summarized. If some image quality can be lost, these compression algorithms represent similar colors as the same color. Another technique for data compression is to transmit only those pixels that changed from the previous screen. Combined, these techniques can reduce the image to as little as 5 percent of its original size, depending on the complexity and stability of the image. This alone reduces the original full-motion video signal to five megabits/second. The image in the previous example, which required 5 Mb/second without data compression, can therefore be transmitted in as little as 384 kb/second, which is one-quarter of a T1 circuit.

The final strategy is to reduce the number of frames per second. In past years, a similar technique was known as slow-scan television. But slow-scan was an analog process that usually displayed a slowly moving scan line across the image. In modern systems, a frame is displayed instantly (or within the NTSC 30th of a second), and then frozen on the screen until it can be completely updated. Systems usually use a slower frame rate as a last resort when other compression methods don't work. Framing rate can be adjusted on the fly, making it a good technique for stochastic packet-switch networks, where network performance is not predictable. Generally, framing rates down to 15 frames/second are not noticeable, except when the camera is panning quickly. Framing rates as slow as 8 to 10 times per second are still usable for traffic surveillance. For video teleconferencing, framing rates as slow as four or five times a second are typical, and such a rate reduces the image- and data-compressed example above to 56 kb per second, obtainable over telephone lines.

The standards for digital video compression are in continuous development by the Motion Picture Expert Group, a working group of the International Standards Organization. Compressed images conforming to their standards are called by their acronym, MPEG.

Most digital traffic surveillance systems, however, have been installed using proprietary compression schemes, which usually exceed the performance of the public standards.

The advantage to digital video is that it shares the same communications path with data. If trunkline services are leased, for example, this advantage is crucial. Excellent video can be ob-

tained from compression to T1 or dual-T1 data rates, and these lines can be leased from telephone companies routinely.

The disadvantage is the cost. In systems where the agency is building its own infrastructure, a digital video systems can represent a huge investment compared to an analog system over the same media. The example from the analog video section showed that the video link could be established with four routine and relatively inexpensive pieces of equipment. Actually, in an analog system, each video signal requires a modulator and demodulator, and all the video signals in a group can share the multiplexors. In a digital system, each video image must first be digitized and compressed, using uses a coder-decoder, or codec. The resulting signal is then piped into the digital and SONET transmission equipment, which at each hub consists at least of a communications service unit (CSU), a data service unit (DSU), a T1 modem, and an OC frame. The cost and complexity of this equipment is significantly greater than analog equipment.

### **Selection Techniques**

Communications systems are complicated. But an understanding of their architecture and technologies can have a huge impact on the cost of a system. About two-thirds of the construction budget for most modern traffic-management systems is spent on communications.

The first step is to determine the requirements. Two questions are critical:

1. Will the system transmit video images?
2. Does the system require mandatory real-time control communications?

If the system will transmit video images, then several more questions must be answered:

- 1a. Will the video be transmitted over leased service?
- 1b. What image quality is required?
- 1c. What is the maintenance capability of the agency?

If leased service is a must, then digital video is probably the only option. If not, then the higher the image-quality requirements, the more cost-effective analog becomes. Limited maintenance capabilities also suggest the simpler analog equipment.

If the system does not require mandatory real-time communications (see the discussion on system architecture in Chapter 3, the reliability of the media can be reduced significantly and wireless media can be considered. Also, systems that don't require real-time control can take advantage of the greater flexibility and interoperability offered by nondeterministic shared-channel networks that use packet-switching techniques. These systems are readily available for all communications media.

If the system does require mandatory real-time communications, then the selection is not so easy. A shared-channel system is still possible, though the system will have to impose the necessary discipline on the communications messages to ensure deterministic performance. Generally, systems requiring

real-time control will lean toward the deterministic private-channel or circuit-switched systems (Appendix J presents an analysis for selecting a communication medium for a Traffic Signal Control System).

### **The National Transportation Communications for ITS Protocol (NTCIP)**

The NTCIP is an ongoing project resulting from a collaboration of the user community and the National Electrical Manufacturer's Association (NEMA). The general trend in systems towards more open architecture has driven the NTCIP's efforts to resolve the problem of non-interchangeability within systems. During the course of the work, however, the NTCIP framers realized the need to expand the scope to cover different kinds of devices that must coexist on and share communications networks, and they expanded the scope of the NTCIP to include most communications links between the control center and roadside systems components of ITS.

The NTCIP seeks to provide two fundamental capabilities: interchangeability and interoperability. Interoperability refers to the ability for dissimilar devices to share a single communications channel. Interchangeability has been defined as the ability to replace a device from one manufacturer with a similar device from another manufacturer without damaging the operation of the system.

#### *a. Interoperability*

Interoperability requires that messages to all devices be defined according to the same rules. In effect, they must use the same protocols at each layer of the ISO-OSI model. As such, all messages must have recognizable wrappings to allow consistent traversal of the communications network.

1. *Protocol Profiles* - A key concern of the NTCIP to accommodate the different kinds of system architectures that are available in the industry. The initial efforts defined communications for distributed systems between master controllers and local signal controllers. These messages must be able to be polled at a reasonable frequency on existing low-speed networks. The focus of the initial work was therefore on efficiency.

But this approach did not accommodate the needs of higher-level communications (e.g. between central and master), nor did it provide a reasonable facility for passing messages along complex networks. The work was therefore divided into two parts: a higher class that would provide logical addressing and a lower class that would only provide physical addressing. Software would be required to unwrap the more complete higher protocol to pass it along a lower network. In a distributed system, this boundary would be handled by a communications hub computer or master controller. In a centralized system, the boundary software would reside in the central computer (at least, theoretically). So the idea of different grades of protocol profiles grew from the need to support diverse system architectures. Since those early discussions, the number of profiles has grown to address different needs within systems. These profiles are:

*Class B*, for communications over twisted-pair copper wire at low speeds. Includes physical addressing only, and can only communicate between a single master device and multi-dropped slave devices.

*Class A*, for communications over higher-performance networks. Provides logical addressing and routing, and can therefore traverse complex networks. It does not provide transmission control,

such as guaranteed delivery and sequential ordering, beyond packetizing. The overhead required for logical addressing and routing renders it unsuitable for low-speed networks, unless a slow polling frequency can be tolerated.

*Class C*, for communications requiring transmission control. In addition to the Class A features, Class C messages also include sequential ordering and guaranteed delivery. Class C also provides a file-transfer service.

*Class E*, currently being conceptualized to cover communications between control centers. At present, the likely result is two profiles. The more efficient profile will provide conventional services and defined message content. The more advanced protocol might use an object-oriented approach, where message content can be defined as part of the message, allowing modifications to the system without changing the software at all control centers.

Only the Class B protocol has been codified into a standard at present.

In the discussion of layered protocols, we defined a profile as a definition of the component protocols across all layers. The protocol definitions for the first three classes are as follows:

Table 2.10.3. Protocol Definitions.

Layer	Class A	Class B	Class C
Application	Small Network Management Protocol (SNMP, part of TCP/IP)  Simple Transportation Management Protocol (STMP, defined in NTCIP)	SNMP  STMP	SNMP  STMP  File Transfer Protocol (FTP, part of TCP/IP)
Presentation	Not Used	Not Used	Not Used
Session	Not Used	Not Used	Not Used
Transport	User Datagram Protocol (UDP, part of TCP/IP)	Not Used	Transmission Control Protocol (the TCP in TCP/IP)
Network	Internet Protocol (IP)	Not Allowed	IP
Data Link	Asynchronous High-speed Data Link Control (HDLC, an industry standard)	HDLC	HDLC
Physical	EIA-232, for serial communications  Bell 202, for 1200-bit/second modems	EIA-232  Bell 202	EIA-232  Bell 202

### b. *Interchangeability*

Interchangeability allows similar devices from different makers to be interchanged on the network without damaging the operation. Interchangeability is more difficult to achieve than interoperability, because the functionality of the system is bound up in the concept. Interchangeability requires the common definition of message content.

Once the messages are defined, they can be transmitted using any appropriate profile of protocols. Consequently, the NTCIP is a definition of protocol profiles, which defines architectural issues, and message object, which defines functionality.

To provide a flexible way to keep up with messages, and to avoid standing in the way of new innovations, the NTCIP employs an object-oriented message structure. Each message is defined by its content, and the devices then interpret what must be done with that object. The objects are defined in a hierarchy, which provides good organization and consistency. For example, all the signal phase timings are part of the phase branch of the tree. The phase branch is part of the actuated signal controller tree, and so on.

Objects are defined for each type of device that will communicate using NTCIP. Ultimately, the list of devices will encompass all roadside systems. Initially, the NTCIP Steering Committee is committed to developing objects for the following devices:

#### Actuated Signal Controllers

- Dynamic Message Signs
- Environmental Sensors
- Camera Controllers (not video images)
- Hub Computers
- Ramp Meters

The first two of these are now complete, and work is underway on the remaining devices.

Within systems, NTCIP can be implemented in a variety of ways. The most critical application resides in the link from the field device to a hub computer or central computer. In these links, messages share the most expensive part of the communications infrastructure, and also the most vulnerable.

To assist the user with applications of the NTCIP, the Steering Committee is preparing a series of tutorials, known as the NTCIP Guides. These will be published by the Steering Committee about the time the first NTCIP-compliant devices will be ready.

User agencies will need a means of testing devices for compliance, and also tools for troubleshooting NTCIP networks. The FHWA is currently preparing public-domain software, known as the NTCIP Exerciser, to provide these services for users. The software will be available without restriction, both in source and compiled versions for use on desktop computers. The Exerciser will allow testing agencies

to send and receive messages from devices that claim to be NTCIP-compliant. The content and semantics of a message will be interpreted by the Exerciser. The Exerciser will also provide a macro-programming capability that will allow users to develop test suites customized for the device in question.

### c. *Standards Development*

The NTCIP represents a joint effort of users and manufacturers. Currently, the NTCIP Steering Committee consists of six representatives each from the Institute of Transportation Engineers (ITE), the American Association of State Highway and Transportation Officials (AASHTO), and NEMA, plus ex officio members from the Federal Highway Administration.

The process for standards preparation and approval meets the requirements for each of the constituent organizations, and is therefore quite complicated. In general, however, the Steering Committee, upon recognizing a problem that needs to be solved, establishes a working group to gather interested parties and create a draft. The process is guided by the Steering Committee, who recommends the resulting draft standard for experimental implementation. When ready, the final draft standard will be formally reviewed through the three organizations, and accepted as a joint standard.

Many working groups are working on various aspects of the protocols. Some groups are quite simple and relatively informal. They may consist of manufacturer's representatives and a few interested users. Working groups can be quite elaborate. For example, the largest working group of the NTCIP Steering Committee is the Transit Communications for ITS Protocol, or TCIP, committee. The TCIP includes dozens of representatives of users groups, system designers, and manufacturers.

The NTCIP development effort is an open process that encourage active participation from all interested parties. More information about the NTCIP and the development process can be obtained from the official NTCIP Home Page on the Internet's World-Wide Web. The address is <http://www.fhwatml.com/ntcip/>.

### **The Transit Communications Interface Protocol (TCIP)**

The Transit Communications Interface Protocol (TCIP) is a family of standards now under development that will focus on data interfaces. TCIP will enable electronic data exchange among transit agency departments, traffic and transportation management entities and Information Service Providers. The TCIP was initiated in November, 1996. Funded by the U.S. Department of Transportation's Joint Program Office for Intelligent Transportation Systems, and developed by the Institute for Transportation Engineers (ITE), the TCIP effort is a one-year standards development effort designed to provide the interface structures that will allow disparate transit components and organizations to exchange data.

As specified in the Project Plan: The project will define the information and information transfer requirements among public transportation vehicles, the Transit Management Center, other transit facilities, and other ITS centers; develop physical and data link requirements; develop required message sets; and establish liaison and coordinate between ITE and other Standard Development Organizations (SDOs) on the development of related standards. It will provide the leadership to coordinate, develop, and deploy a comprehensive set of transit communication interface protocols (TCIP) that allows effective and efficient

exchange of data used for ITS user services and transit operations, maintenance, customer information, planning and management functions. The scope provides for interfaces among transit applications to communicate data among transit departments, other operating entities such as emergency response services and regional traffic management centers. The results of this effort will be to a set of provisional standards for TCIP.

The ambitious schedule will be coordinated with the TCIP Technical Working Group (TWG) to ensure industry review at every stage (Paula Okunieff, et al, February 24, 1997).

### 2.10.2 TECHNOLOGIES

The following technologies are used for communication of freeway systems.

#### **Twisted-Pair Cable**

Twisted-pair cable is individually insulated copper wires twisted into pairs, and combined into a shielded cable. Agency-owned twisted-pair cable is widely used for the low-speed transmission of traffic control data (e.g., between hubs and field elements), with the network configured with between 8 and 16 field drops on each two-pair (4-wire) channel. The exact number of drops depends on the amount of data to be transferred between central hub and the field locations, and the rate of transfer.

Video transmission is possible on twisted-pair cable but requires repeaters every 3,000 feet to 3.6 miles, depending on the wire gauge. Typical data rates are 1,200 bps. Voice can be carried on a single pair. Bandwidths range from 300 Hz to 3,000 Hz.

Installation of twisted-pair cable can be accomplished by: underground in conduit; underground by direct burial; aerial utilizing existing or new utility poles. The method used will have a significant impact on cost.

Advantages:

- Represents a low cost form of transmission.
- Easy to splice.
- No specific interface equipment required.
- Electrical characteristics are favorable to basic analog transmission.

Disadvantages:

- Data cable splicing is not recommended.
- There is a bandwidth limitation in that twisted pair cable tends to attenuate high-frequency electrical signals, thereby limiting the ability to transmit digital information at high rates.

- Bandwidth limitation prevents transmission of live television images, though recent developments permit transmission of slow-scan television. (Prototype equipment is available for transmission of full-motion television over twisted-pair copper wire).
- No protection from Electromechanical Interference or Radio Frequency Interference (EMI/RFI) when using twisted-pair cabling facilities for video.
- Low security.

### **Coaxial Cable**

Coaxial cable consists of a center conductor surrounded by a dielectric material (insulator) and enclosed within a grounded conductive outer shield. The center conductor is typically copper-clad aluminum; the dielectric may be a solid (foamed polyethylene) or gas; and the outer conductive shield is typically aluminum and may be braided metal fabric, a corrugated semi-rigid metal, or a rigid metal tube. There is also an outer jacket consisting of low-density, high-molecular-weight polyethylene. This construction significantly reduces the effect of external radiated or induced noise, particularly at higher frequencies. Very large bandwidths (5MHz to 550MHz) may be transmitted over coax. This bandwidth can be divided to handle more than 40 channels on one cable, with additional capacity to handle multiplexed data channels. Coax is especially flexible, supporting a wide range of data rates can be supported in coax (from 1200bps to 10Mbps). The cable can also support shared use in integrated systems.

Advantages:

- Because of its physical structure, coaxial cable is more immune to electromagnetic interferences and has a much higher bandwidth than twisted-pair cable.
- Minimal signal losses.
- Low signal leakage.
- Higher bandwidth allows for transmission of video signals (cable television using coaxial cable can transmit as many as 75 independent video signals) and for the transmission of digital data at very high rates.
- Bandwidth of coaxial cable permits theoretical transmission rates as high as 700 million bps.
- A single high-speed channel can serve an entire system (many channels of information can be multiplexed onto a single space). This is an advantage in urban areas where cable sizes are limited by available conduit space.

Disadvantages:

- Dedicated systems need repeaters spaced at 1-km increments.
- Splice connections are susceptible to noise and transient problems.

- Coaxial cable can not be spliced together by manual strip and twist method. Inherent nature of cable and the importance of conductor alignment make the coaxial cable much more difficult to splice.
- Lower communication reliability than fiber optic.
- Higher maintenance and adjustment effort required compared to fiber optic.
- Low security.

### **Leased Telephone**

The most common type of leased telephone company (Telco) channel is the 3002 conditioned analog channel, which transmits voice-grade data (300 Hz to 3,000 Hz) and can be used as a twisted-pair channel with multiple controllers operating on a TDM channel. Typical available leased circuits include voice-grade/data grade analog (1,200 bps), digital (56 kbps), T-carrier and SONET. Data rates achievable with 3002 are typically limited to 1,200 bps. Voice frequency FSK modems are required and can be configured for either full- or half-duplex operation. Leased telephone lines are frequently used for traffic surveillance and control system applications. These applications include interconnection between a central computer and field equipment such as traffic-signal controllers, DMSs, freeway ramp meters, traffic counting equipment, etc.

Advantages:

- The network is already in place.
- Reliable communications solution in that there is a grid redundancy element due to the general coverage of the carrier's network.
- No responsibility on the lessee to maintain the required infrastructure. That responsibility lies with the telephone company.
- Capital costs are very low, involving connection from access points to cabinets and central facility.
- Appropriate where a high volume of voice or data communication is required between separate facilities.
- Can be used for temporary applications.
- Used where other forms of communication technologies are not cost-effective.

Disadvantages:

- High lease costs, always tied to general service rate increases.
- Less control of the communication network; requirements and limitations on leased channel may impede use.
- No flexibility in the design of the communication network; constrained by telephone company cable network.

- Not a cost-effective communication medium for distant communications requiring multiple central offices.

### **Leased Cable Television (CATV)**

Because CATV (cable television networks) are usually implemented on coaxial cable, the characteristics of CATV channels are very similar to those of dedicated coaxial cable systems. Two major differences: 1) the available bandwidth on CATV, which is a function of the network loading and the franchise agreements, may be limited; and 2) not all CATV systems are designed for bidirectional communications. A potential problem with CATV is that the existing area of coverage and the network layout may not coincide with the traffic signal density.

Advantages:

- A single 6-MHz channel is adequate for data transmission.
- The network is already in place.
- Design efforts and initial installation costs are very low.
- Franchise agreement may provide for government use of CATV cable and bandwidth at reduced rates or free, reducing recurring costs.
- Second separate coax institutional network (I-net) may exist for the express purpose of providing bidirectional services to commercial subscribers.
- I-nets generally provide good levels of service to subscribers.

Disadvantages:

- Most CATV networks designed and installed with emphasis on downstream transmission of video signals to cable subscribers.
- Video channels take up most available bandwidth.
- Bandwidth available to traffic control may be very narrow, ranging from a single 6-MHz channel to four or five channels.
- Single 6-MHz channel does not support full-motion video transmission in addition to data communications.
- Frequencies of available channels are often most susceptible to noise and interference.
- Quality of video signal required for CATV considerably less than required for data transmission.
- CATV subscribers sometimes concentrated in residential areas.
- Service to Central Business district and industrial areas sparse or nonexistent.

- Area of coverage and network layout may not coincide with traffic-signal and other equipment locations.
- Traffic-control system may have to compete with other public I-net users for more desirable channels.

### Fiber Optics

Fiber optic technology consists of a nonmetallic conductor that guides light rays. The glass fibers are coated with a cladding of glass or plastic having a refractive index different from the glass in the fibers. The difference in refractive indices causes light waves striking the cladding layer below the critical angle of incidence to be reflected back into the core section. In this way, light rays are propagated to the other end of the fiber with low intensity loss.

The use of fiber optics requires a dedicated, owned communications network. This requires right-of-way and conduit throughout the network. Right-of-way is usually the limiting factor for private companies, but not for the State. The cost of installing and maintaining conduit, however, can be significant. The information-carrying capacity, or bandwidth, of fiber is related to the spreading or dispersion of the light pulses. Pulse dispersion is a function of both pulse width (data rate) and distance. The capacity of fiber-optic systems is expressed in Mbps-km — the product of the data rate and the repeater spacing. Capacities of up to 500,000 Mbps-km (4Gbps at 117 km) have been achieved. The maximum repeater spacing in fiber-optic networks is a function of both the bandwidth of the transmitted signal and the signal attenuation.

One of the most common, easiest methods for providing multiple channels is to simply install a multiple-fiber cable, and use a separate fiber for each channel. Because fibers are so small (more than 200 bare fibers can be placed side by side in the space of an inch) a cable of 12 or even 50 fibers is smaller than a single twisted-pair copper cable. In addition, due to the high cost of sophisticated multiplexing electronics, installing multiple fibers can be more cost-effective.

Fiber is immune to electrical interference. Since the signal is optical in nature and there is no metallic medium, fiber-optic transmissions are not affected by cross talk, ground loops, or ingress. Fiber is essentially a point-to-point technology — in order to serve multiple drops on a fiber circuit, it is necessary to use a drop and insert modem. One application is trunking where several low-speed channels are multiplexed into a high-speed data channel and transmitted over a single fiber from a central point to a remote distribution point. At the remote location, the multiplexed signals are demultiplexed into low-speed channels and then transmitted over twisted-pair cable to the field drops (controllers). Similarly, the responses from the controllers are gathered at the remote distribution location, multiplexed together, and transmitted back over another fiber to the control center, where they are demultiplexed and read by the computer.

Fiber-optic cable may be installed in conduit, direct burial, or aerial applications.

Advantages:

- Large capacity.

- The capability of expansion is practically unlimited, as fiber bandwidth is very large.
- Immunity to electromagnetic and radio-frequency interference.
- High integrity for data transmission.
- Emits no radiation. Represents a highly secure means of communication, because it is difficult to tap a fiber tube without detection of the resulting signal loss.
- A small, flexible (small bending size and radius), and lightweight cable.
- Safety in hazardous environments.

#### Disadvantages:

- Designing a fiber-optic network requires substantial engineering effort, due to complexity of networks, light distribution characteristics and mediums, and other factors.

### **Microwave**

Microwave is an air-path medium; it does not require a physical connection between the transmitter and receiver. Signals are radiated from an antenna, propagating through the atmosphere along a line-of-sight path. Frequencies used must be unique in that area and direction to prevent interference from other microwave transmissions. Because of this constraint, microwave frequencies are licensed by the FCC. It is, therefore, very difficult to obtain microwave frequency allocations in crowded urban areas. When frequencies are available, they are usually in the higher frequency bands (18 and 23 GHz), which have reduced transmission distances.

Configurations for microwave communications are typically point to point or point to multipoint, depending on the requirements of the network. The point-to-point network is one where the transmitter propagates the microwave signal to a receiver located some distance away (0.5 miles to 40 miles, depending on the frequency utilized). The point-to-multipoint system requires one transmitter propagating the microwave signal to receivers at multiple sites in the field.

The utilization of microwave communications falls into two major categories. The first is an analog microwave system in which the method of modulation is typically FM. This type of system is utilized for the transmission of video images (CCTV) which provides a full-motion image. The second is a digital microwave system in which the method of modulation is typically TDM/TDMA (Time Division Multiplexing/Time Division Multiple Access). This is an all-digital system which is used for data and voice. Video may be transmitted through use of digital video compression techniques.

Frequencies are in the range above 1 GHz (1,000 MHz). Microwave frequencies are attenuated by the atmosphere, with attenuation increasing with frequency. Rain and fog can be detrimental to microwave transmissions. The southern California region's mild climate makes it ideal for microwave, while climates of the Pacific Northwest, the East Coast, and Midwest are considered more difficult in the design of a microwave communications link.

Since microwave is an air-path medium, it is regulated by FCC and the available bandwidth is a function of how much bandwidth has been allocated in the area and how much is still available. Maximum data rate is a function of the available bandwidth — 10 Mbps data stream can be transmitted over a channel only 7 MHz wide.

Microwave can be used like fiber in trunking applications for data. In one application, signals along an arterial, located 10 to 15 miles from the control center are interconnected into three full-duplex data channels via twisted-pair cable. These data are multiplexed together and transmitted to the control center over a microwave link. In another application, microwave is used for video transmissions across a large body of water.

Advantages:

- Useful as a point-to-point trunk.
- Can transmit data and a limited number of full-motion video channels.
- Can control groups of traffic-control devices.
- Can use both analog and digital transmission.

Disadvantages:

- Requires line-of-sight path.
- In most cases, requires FCC license.
- Limited channel availability.
- May offer little choice of operating frequency its very difficult to obtain microwave frequency allocations in crowded urban areas.
- Possible interference due to rain, snow, and atmospheric conditions.
- May require antenna tower.
- Available bandwidth usually limited; this is a function of how much has been allocated in the area.

### **Satellite**

Satellite is similar to microwave technology in that both use some of the same frequencies for transmission through the atmosphere. With satellite, however, instead of utilizing a line-of-sight transmission path, the signal is directed to a satellite transponder. Very Small Aperture Terminal (VSAT) satellite systems use 14 GHz and the downlink (i.e., transmissions from the satellite to the earth) use 12 GHz. The satellites themselves are in a geosynchronous orbit above the Earth's equator, thereby appearing stationary in the sky and providing 24-hour-a-day coverage.

In addition to the satellite links, a central hub element and earth stations are also required to provide a VSAT-based communications network. The "hub" may be located at a traffic management center, and

serves as the focal point for communications between the remote sites and the control center. The “earth station” is the communications component of a remote site. The earth station consists of an antenna, an RF transceiver mounted on the antenna, and a digital interface unit. For video applications, the digital interface unit is connected to the camera’s central receiver (for pan, tilt, zoom) and a codec unit.

In general, the VSAT medium is ideal for long-distance communication links (e.g., communications between the Caltrans District 12 Traffic Operations Center and Sacramento) since the cost of leasing VSAT channels is independent of distance. In general, VSAT becomes cost-effective relative to leased lines at a transmission distance somewhere in the range of about 300 miles.

Since VSAT lease costs are a function of the percentage of time that the inroute/outroute link is actually utilized, this technology may also have cost-effective applications in a localized communications network where transmissions are required only on an as-needed basis, such as for CCTV that is only used as needed for incident verification.

Another potential use of VSAT is for communications where a dedicated cable network and/or leased telephone circuits are not readily available, or where a surveillance camera needs to be mobile. For example, VSAT may be used in conjunction with the maintenance and protection of traffic during roadway construction, and during major incidents in areas where CCTV is not available, but real-time video surveillance would be a great asset.

#### Advantages:

- Cost of circuits is independent of their length.
- Cost-effective for long-haul communication links.
- Downlink signals can be received over a wide area.
- Cost-effective for point-to-multipoint distribution applications.
- Uplink signals can originate from a wide area.
- Flexibility for temporary or mobile applications.
- Provides option to links that cannot be achieved by earth-based communications media.
- The high-altitude satellites avoid various earth-level interferences.

#### Disadvantages:

- Not proven cost-effective for local communications.
- Limited number of service providers.
- Channel-leasing costs subject to increases.

- Most transportation agencies have no need for the frequent long-distance transmission capabilities of satellite technology.

### **Optical Infrared**

An air-path medium, infrared signals are transmitted from an infrared LED or laser diode (LD)S through the atmosphere along an unobstructed line-of-sight path to the destination. The mode of transmission is similar to the microwave but at a much higher frequency — higher than 300 GHz. Because of its narrow beamwidth and low-power infrared, transmission is not regulated by the FCC. The bandwidth, as with fiber optics, and the maximum data rate of infrared links is dependent on distance. At short distances (less than 1 mile), an infrared link can support full-motion video. At distances up to 10 miles, infrared links can support a 9.6 kbps serial data channel. Infrared is essentially point-to-point transmission, like microwave. One application might be where many channels of data are multiplexed into a single high-speed channel for transmission to a distribution site. Another application might be the transmission of CCTV video signals from the cameras to the control center.

Advantages:

- Like microwave, it is a point-to-point transmission.
- Higher frequencies than the microwave (higher than 300 GHz).
- Transmission not regulated by FCC.

Disadvantages:

- Requires line-of-sight path.
- Severe weather conditions such as fog, smog, or smoke (but not rain) can cause degradation of the infrared transmission.

### **Radio**

Radio technology transmits radio frequency electromagnetic energy — not contained within a waveguide (air-path) through space. Radio frequencies can range from the hundreds of kHz to 1 GHz, and are regulated by FCC. The most likely bands in traffic control are the 928/952 MHz bands, the 450/512 MHz bands, and the cellular radio bands. These frequencies are reserved for point-to-multipoint, master-to-remote, duplex data communications. The total bandwidth (number of channels) is dependent on the number of users and how many channels are already allocated.

Advantages:

- Can provide voice communications to highway maintenance vehicles.
- Can propagate into built-up areas and buildings.
- Can support 9,600 baud data rate.

- Can prove cost-effective depending on application.

Disadvantages:

- Terrain may limit range.
- Limited channel availability in urban areas.
- Because of the narrow channel width, radio channels are not usually multiplexed.
- Requires antenna at each controller site.
- Turnaround time excessive for some applications.
- Service reliability may limit use in some applications.
- Regulated by FCC.

### **Cellular Radio**

A cellular radio is technique for frequency reuse in a large radio communications system. Cellular radio is based on the concept of "cells" which are 2 to 20 miles across. The center of each cell is a control radio that handles the network management functions, including the assignment of frequency sub-channels. A radio requests a frequency over a control channel and one is assigned by the cellular control system. Frequencies are in the 825 to 890 MHz band.

Instead of allocating a channel for a single user over an entire area, a large group of frequencies is divided into cells. This allows nonadjacent cells within the system to use the same frequency channel without interference, i.e., frequency reuse. One application might be a communication between a portable remote terminal and the central computer. Another application might be dial-up connection between the central microcomputer and the on-street master in a distributed system.

Advantages:

- Allows frequencies to be re-used without interference.
- Frequency reuse tremendously multiplies the number of channels available.
- May prove cost-effective for infrequent communications.
- Eliminates need to connect to a telephone company service.
- Effective for controlling portable DMSs.
- May prove effective for temporary installations.
- Network covers approximately 93 percent of the U.S. population.

Disadvantages:

- Radio transmission of data at low power levels is susceptible to noise.
- Airtime cost excessive for continuous communication service.
- Actual data throughput reduced due to protocol overhead.
- Remote areas may not have service.

### **Spread-Spectrum Radio**

In spread-spectrum radio transmission, the signal occupies a bandwidth in excess of the minimum necessary to send the information. The band spread is accomplished by means of a code that is independent of the data, and a synchronized reception with the code at the receiver is used for subsequent data recovery. With spread-spectrum radio, the entire band (i.e., 902 to 928 MHz, as designated by FCC) is available for use by all users. This is accomplished by “spreading” the signal over the entire 26-MHz spectrum (in the 902 to 928 MHz band) and requires that the receiver “lock-on” to the transmitted signal, which adds to the efficient use of the available spectrum. Due to the spreading of the signal over a wide frequency range, electromagnetic noise (interference typically generated at a very narrow band width) has less effect on the signal integrity. This system is designed to replace the physical medium links in most closed-loop systems — consisting of half-duplex, 1,200 baud, TDM/FSK communications.

Advantages:

- Flexible installation.
- Does not require cable installation and maintenance.
- Does not require FCC channel, use approval in the 902 to 928 MHz band.
- Works extremely well in a high-noise environment.
- Currently in use for many industrial process control applications.
- Uses low transmitter power.
- Can be used in a mixed system of wired or radio interconnected controllers.
- No land-line interconnect required.
- Relatively low equipment cost.
- Potential for broad range of traffic control-system applications.

Disadvantages:

- New technology for traffic-control and -surveillance applications.

- Power restrictions limit the transmitter output to approximately 10 miles.
- Higher bandwidth than radio fixed-frequency transceivers.
- Requires external antenna and cable.
- Unprotected channel space.

### 2.10.3 CASE STUDIES

TransGuide - From the TransGuide ITS Design Report ([transguidewwwserver.datasys.swri.edu:80/articles/designreport.html](http://transguidewwwserver.datasys.swri.edu:80/articles/designreport.html) )

#### Communications System

The communications system ties all of the other pieces of equipment in the system together. One of the basic attributes of a traffic information system is its geographically dispersed nature. The sensors used for detection, verification and characterization of incidents must be distributed throughout the road system, and the information produced by the traffic management center must be distributed to the drivers on the roads to be useful. There must be a system for collecting the raw data used to generate traffic information and for disseminating that information to the drivers.

Several architectures were considered as the basis of the TransGuide system. Some of the architectures consist of a single physical medium. Others contain a variety of media. Selection of the architecture, the physical media and the mechanism for switching the control and video was based on the fact that those attributes are interrelated.

#### Communications System Architecture

Several different types of communications system architectures are used in traffic control systems, and several were investigated as part of the specification task for the TransGuide system. Primary considerations included the system maintainability (including the availability of network management capabilities), equipment requirements, environmental requirements, reliability, and supportability.

Factors investigated included the overall architectures, the level of performance, and the impact of funding mechanisms. Traffic control systems frequently contain several types of communications media; one to handle the data and control functions, one to handle the video signals, and another to handle future ITS applications. These systems typically contain a T1 network, a multimode fiber optic cable, and a single mode fiber optic cable. The T1 network is used for voice and data transmission, the multimode fiber optic cable is used for transmitting a modulated version of the analog video signal, and the single mode fiber optic cable is intended for future ITS requirements.

The use of three separate systems significantly increases the problems associated with managing the networks. There are no off-the-shelf network management systems available to manage three disparate networks. In fact, separate off-the-shelf network management systems capable of managing all three types of networks independently were not readily available while TxDOT was specifying the

TransGuide system. The complexity of maintaining more than one type of network, and the additional cabling and connecting required for separate networks, exacerbates the management problem.

One communications system considered for distributing video signals included modulation and demodulation equipment for each camera, one on each end of the fiber optic cable. In this configuration, providing the video signal to any of the displays would require a large analog video switch capable of switching any one of the cameras (eventually up to 1500) to each of the displays in a non-blocking fashion. The system requirement to be able to view any camera on any of the CRTs (simultaneously, if desired) places this architecture at a severe disadvantage. A communications system in which each camera is hardwired to a specific CRT does not require a large analog switching capability. However, it does not meet the system requirements.

Another communications system choice was to use frequency division multiplexing of the video signals onto a common set of cables. Initially, a single cable could handle all of the cameras, but eventually three to ten cables would be required, depending on the number of channels multiplexed onto each cable. This communications system would operate much like a cable television system, minimizing the overall amount of cable installed. It would also primarily use off-the-shelf components. However, some parts of the system would probably not be available off-the-shelf. The system would also require either separate data and voice networks or would require custom equipment to transmit voice and data signals over the cable.

The communications system eventually selected includes the use of dual single mode fiber optic cables from fiber hubs located strategically throughout the system. The cables are separately routed. An automatic fail-over mechanism switches to the back-up cable in the event of a failure of the primary cable or drive laser. There are several significant advantages to this communications architecture. The digital fiber optic system is compatible with the use of Synchronous Optical Network (SONET), a telephone industry communications standard. The use of the same single mode fiber optic cable used by telephone companies and the use of SONET indicate that long-term support will be available. The selection of communications equipment used by telephone companies also means that there are off-the-shelf network management systems available to control the network, which means reduced maintenance costs and increased system reliability. The selection of the same fiber optic cable, interfaces, and protocols used by telephone companies also means that equipment is standard and available off-the-shelf as well.

The availability of environmentally hardened field equipment for this communications architecture also allowed system designers to minimize development of special equipment and operational maintenance costs while maximizing the reliability of the system. One of the driving factors in the selection of the OC3 based communications system was the availability of environmentally hardened (primarily high temperature) equipment to perform the functions required in the fiber hubs. The fiber hubs must be located near the cameras themselves and are therefore geographically distributed. The use of equipment that is not environmentally hardened in non-environmentally controlled fiber hub cabinets is incongruent with the reliability and availability goals of the TransGuide system. If the communications system components required for the fiber hubs were not available in temperature hardened versions, the fiber hubs distributed throughout the city would have to be environmentally controlled, significantly increasing both the initial cost of the system and ongoing operation and maintenance costs.

Diversity can be provided in fiber optic systems by two different physical connection architectures. One architecture is defined as a ring, and redundancy in a ring is based on multiple routes between nodes. A second architecture is defined as a star, and redundancy in a star is based on route diversity. The selection of the best communications architecture for the TransGuide system was based on the unique attributes of the TransGuide communications system. In the ring architecture, each node is connected to two other nodes to form a logical ring. Each node can communicate in either direction around the ring. Therefore, a ring configuration does not allow a single failure (such as a cut conduit) in any one place to isolate any node on the ring from any other node. Most telephone company systems are built using the ring architecture. A system based on the ring architecture requires communications equipment at each node capable of handling all communications through that node. Each node in a telephone company communications system generally communicates in a statistically distributed manner to each of the other nodes in the system.

Random node-to-node communications are not a requirement of the TransGuide communications system. Each TransGuide node communicates only with the TOC node. In a ring architecture, all communications from each node would need to be capable of reaching the TOC node in either direction. Each node would require the ability to transmit all of the communications from most of the other nodes in the ring. To implement such a ring architecture in the TransGuide system, multiple rings would have had to be configured using OC24 or OC48 equipment in the fiber hubs. Neither OC24 nor OC48 equipment is available in environmentally hardened versions.

The Irvine's Field Operational Test uses Asynchronous Transfer Mode (ATM) communications technology to carry out the real-time bi-directional exchange of data and video. The Field operational Test will integrate the Caltrans TMC and Irvine's Arterial Response Plan system module using MIST via an Ethernet connection over a fiber optic link. The infrastructure includes 28 type 2070 ATCs, additional OPAC loops and five arterial DMSs at strategic locations. The 2070s contain traffic control firmware which exchanges real-time data with the OPAC algorithm through the Data Exchange Module. The key ATM attribute is its capability to support multimedia and integrate such services along with data over a signal-type transmission method (Karna et al, August/September 1996)

## **CHART**

A hybrid approach of both building and leasing telecommunication facilities was taken by the State of Maryland and their CHART system. They did an analysis of the communications needs and an assessment of the build vs lease tradeoff. The study took in some 600 miles of highway with over 1000 field devices, including 200 CCTV cameras, loops and radar detectors within the CHART system as well as telecommunication needs to link five local district operations centers, highway maintenance facilities, other district offices and park and ride facilities. With the size of this communication network, the distribution of data was a major issue. They analyzed eight different network configurations taking into consideration also the existing infrastructure which is 100 miles of fiber optic backbone in the Baltimore-Washington metropolitan area. With the amount of video footage they had to supply to various broadcasts, they have implemented compressed video transmission at 384 kilobits over T-1 phone lines throughout the state of Maryland.

The Maryland study revealed that it would be more cost-effective to use wireless communications to link certain field devices than it was to trench with wire. They have Cellular Digital Packet Data network in

place across Maryland (through Bell Atlantic) and can provide a cost-effective method of sending data (Ian Nuttall, August/September 1996).

Some other conclusions drawn from the Maryland study included (Transport Technology Publishing, October 1996):

- Where the density of roadway devices, detectors, cameras, etc., is low, there is no way to economically justify building a network to serve these areas. It was considerable cheaper to use wireless communications to get data from devices to the network nodes than laying cable to the devices in many cases.
- The telecomm industry has changed and is continuing to change rapidly and new technology, already operational in some locations, will greatly reduce the cost of the leased network over the next several years. It was, therefore, recommended that a 3 year lease be chosen vs anticipated 10 or 20 year lease the industry was willing to provide.

## 2.10.4 EVOLVING TECHNOLOGIES

### Video Through Telephone

Telephone provides the most cost effective means of video transmission. There are two main categories for telephone circuits: Dial-up and Leased. Dial-up uses the same lines as those used for everyday voice communication in making a call or sending a fax. One is charged only when using it. Dial-up is mostly used when it is not necessary to view a camera for long periods of time. It is very cost effective when used for short periods of time in viewing an area or sending a video to the traffic control center. Leased telephone lines are used when it is necessary to use video surveillance for long continuous viewing of an area.

### SlowScan Trough Telephone Lines

This technology uses dial-up phone lines to transmit video using scanning techniques. The video is scanned or painted onto the monitor like a snap shot and pictures are refreshed every 25 to 40 seconds. This technology is very efficient and cost effective in traffic management where real-time video is not needed and only snap shots of traffic conditions are required and updated, at approximately one-half to one minute. One application where this technology is used is to view and confirm messages on changeable message signs. Speed of image transmission is depended on the resolution required. For higher resolution images, such as reading a license plate, the speed of transmission is slower than low resolution images.

### SlowScan Trough Cellular Phone

This technology uses cellular phones to transmit slow-scan video. One application where this technology is used is where temporary cameras are needed to gather images or other information that are isolated from the network. Since the use of cellular phones is increasing at an alarming rate, this technology can be used in the future to gather surveillance information from just about anywhere.

## Digital Compressed Video

Where it is infeasible to install cable or lease cost for fiber is prohibitive, compressed video offers an attractive alternative. Transmission of full motion video images is 30 frames per second. Past compressed video techniques, refreshed rates of eight to ten frames per second.. This resulted in the image appearing slightly jerky. However, with new technology, it is now possible to compress the video so that it fits onto the telephone line at full motion. The picture delivered is almost real-time, but to the naked eye and for the purposes of traffic management and surveillance it is sufficient. For instance, to represent one frame of a video in the 0s and 1s of computer language, it takes over a megabyte of memory. Since there are 30 frames in the standard second of video, one would need 33 MB of memory transfer or storage for one second of video. Compression methods can reduce this data at nearly a 200:1 ratio — creating a much more manageable and cost-effective data stream. The process of video compression uses a process that shrinks images so that they occupy less storage space and can be transmitted faster and easier. The bits that define blank spaces and other redundant data are removed and are replaced with a smaller algorithm that represents the removed bits.

Digital compression technology will be the key to multimedia and other emerging video communications. Products and services affected by compression include multimedia personal computers, video conferencing, interactive video training and games, high-definition television (HDTV), digital broadcasting services and many others.

## Laser

This technology is similar to the fiber optic technology where the signal transmitted is converted from an electrical signal to a light wave signal. However, unlike fiber, the transmission is wireless, through space. There is no FCC licensing involved, no right-of-way fees, no cable installation needed, is not susceptible to EMI or RFI interference and has a very high bandwidth. The disadvantage of the laser technology is that it has a very short operating distance, approximately one mile, needs a straight line of site for the laser transmission, and the video quality is affected by rain, fog, and snow.

## SOURCES:

- *ATM Networks in ITS*, August/September 1996, by Kamal Karna, Jack Bailey, and Jim Arnold, COMtrans.
- *Getting to the Bottom Line on Telecommunication Costs*, August/September 1996, by Ian Nuttall, COMtrans.
- *An Overview: Transit Communications Interface Protocols (TCIP)*, February 24, 1997, by Paula Okunieff, Tom Buffkin, Xudong Jia, Karen Levine, Eva Lerner-Lam, and Isaac Takyi, ([www.tcip.org/paper1.html](http://www.tcip.org/paper1.html)).

## 2.11 IN-VEHICLE NAVIGATION SYSTEMS

### 2.11.1 TECHNOLOGIES

In-vehicle navigation systems combine advanced Global Positioning System (GPS) technology for precise navigation, detailed digital moving maps able to track a vehicle's progress in near real-time, and a yellow pages directory of instantly acceptable travel information about restaurants, hotels, entertainment, and shopping (Wilsterman, 1994). The GPS receiver integrated into these systems provides the position of the vehicle on the road in real-time mode, and display that position on a map in a high-resolution color Liquid Crystal Display (LCD) screen. The map, as well as the travel information, is recorded on a CD-ROM disk drive. Wilsterman (1994) describes the operation of these devices as follows:

The driver of the vehicle selects his/her route in real-time while driving. This can be done by entering the destination's name or address, for instance, home or office. The GPS navigation system receives positioning signals from the NavStar network of 24 satellites to pinpoint the vehicle's geographic location, including latitude, longitude, and altitude, and places it on the in-vehicle display map within 100 yards of the car's actual position. The GPS receiver can be set to automatically update the vehicle's position every second, continuously displaying the vehicle's movement in real-time. This makes it easy for the driver to choose a route, evaluate alternatives, and change course in response to road and traffic conditions.

These systems have a built-in zoom function that, just like a video camera, zooms in on a location or pans back, allowing the motorist to reference multiple viewing levels or stages. The zoom illustrates details of individual streets and neighborhoods on its lowest level, to an eagle's-eye view of entire cities, regions, and highway networks on its highest level.

Accessed by a wireless remote, the digital maps are easy to read with clear text and rich graphics that not only identify but also distinguish between interstate highways, State roads, local streets, alleys, parks, landmarks, airports, and natural boundaries. EtakGuide, a product compatible with Sony's navigation technology, uses audio and video prompts, icons, and colorful screen menus to help novices use the system.

Krakiwsky (1996) categorized the navigation systems into five types: autonomous, advisory, fleet management, inventory and portable. The first three types are classified according to their communications capabilities. Autonomous systems have no communications; advisory systems normally have one-way communications — receipt of traffic information, for example; and fleet management systems have two-way communications between the base station and each vehicle. Inventory and portable classes reflect the specialized use and important portability characteristics of some newer navigation systems. The inventory category includes systems related to time- and coordinate-tagging of road-related information for road inventory and other surveillance purposes. The portable category covers products that are the natural extension of portable data assistants (PDAs) — palmtop and laptop computers, for example — into personal navigation assistants (PNAs). The boundaries between these categories have become more blurred as products evolve.

The three major technologies for determining a vehicle's position are: the global positioning system (GPS), dead reckoning and map matching. With selective availability (SA) turned on, accuracy is about 30 meters but can degrade further to about 100 meters. SA is the deliberate degradation of GPS satellite signals by the DOD to withhold real-time capability from undesirable users. Differential GPS (DGPS), a technique used to improve accuracy, offers precision to within one to ten meters, but corrections for the effects of SA must be sent to the vehicle to get this improved accuracy. Using GPS requires a direct line of sight from the receiver to the satellite, so it must be used in conjunction with other positioning methods if vehicles will be traveling in urban canyons (streets surrounded by tall buildings) or tree-covered areas. GPS is a universal, low-cost technology — a few to several hundred dollars -- that is revolutionizing navigation.

Dead reckoning can supplement or replace GPS and uses a selection of speedometers and odometers, a compass, and a gyroscope to determine the direction and distance traveled by a vehicle. The system must be initialized using the coordinates of a known starting point. The sensors' accuracy is fairly high for short periods of time, but they require corrections by GPS or map matching over longer periods to avoid error accumulation, which can reach a few hundred meters during extended travel. Dead reckoning devices can double the cost of a system and make it difficult to move the equipment from one vehicle to another.

Map matching is used to augment the less-than-perfect positioning provided by satellites or dead reckoning. It assumes that the vehicle is on a road contained in its digital map database. When a positioning system provides coordinates that do not exactly match those of a link, a road segment in the digital road map, the map matching algorithm finds the nearest link, and snaps the vehicle onto it. Changes in the vehicle's direction and distance traveled are tracked, and the pattern of travel is compared with the road network in the map database to pinpoint vehicle location.

Navigation and tracking systems, with the exception of autonomous systems, rely on one- or two-way voice and data communication. Several types of communications technologies can be employed, including: conventional radio, cellular systems, trunking, satellites, beacons and signposts, radio data systems (RDS) and paging. Communications devices are the fastest-changing component of tracking and navigation systems. They affect the cost more than almost any other system components. The choice of communications depends on the intended application and user requirements such as reporting rate, throughput capacity, geographic coverage, and cost.

The most commonly used communications system in automatic vehicle location (AVL) designs throughout the world is conventional radio. Equipment may be portable or mobile, the type often dash-mounted in vehicles such as taxis. This system is intended to operate in defined geographical regions; users share a common radio frequency channel and compete for air time.

Cellular's wide acceptance (19 million-plus U.S. users) and national coverage (approximately 90% of the U.S. population and 60% of its geographic regions) have led to significant interest in the AVL marketplace for cellular communications as a means to transmit data over the air. Cellular phones are relatively inexpensive, and airtime costs are decreasing with more competition.

Some satellite-based systems complement existing terrestrial-based systems — cellular and special mobile radio, for example — to fill in uncovered areas.

Beacons and signposts communication devices transmit and receive information by microwaves, infrared light waves, or short distance radio frequencies. They are part of the fixed infrastructure and are typically situated at intersections to exchange information with passing vehicles. This type of system requires a heavy investment in infrastructure.

According to Kelly Tisdell and Edward J. Krakiwsky (1997), more than 400 land navigation, tracking, and positioning systems are on the market today. These systems make up: route planning and optimization, fleet tracking, route guidance and MAYDAY.

### **Route Planning and Optimization**

One of the most popular route-planning and optimization products on the market is Roadshow (Roadshow International, McLean, Virginia). The system uses an algorithm to calculate the least-cost route and allow users to consider variables such as traffic bottlenecks, one-way streets, and rush-hour traffic delays. These variables are specified by the user during calibration based on past experience and current knowledge of the route and are adjusted as needed to respond to changes. A Windows-based personal computer displays customer data and a map showing the best route and allows users to modify that route. Roadshow is fundamentally a stand-alone route-planning system — meaning a PC workstation is dedicated to the task of route management and dispatching — but communications capabilities are an option so dispatchers can track and reroute a fleet of vehicles in real time.

### **Route Guidance**

Navmate a popular route-guidance system, integrates a GPS receiver, navigable electronic map databases, a directional sensor, and a control processor. The driver inputs a destination — an address or point of interest — and the system offers turn-by-turn directions with audio and visual cues. If traffic is stalled by construction or an accident, the driver can abort the current route. The system will then calculate an alternative route that avoids the problem. If a driver makes a wrong turn, the system calls attention to the error and asks if the driver would like to calculate a new route based on the vehicle's current position.

Zexel has licensed Navmate to General Motors, Rockwell International, and Siemens Automotive. Rockwell's PathMaster, a licensed version of Navmate, is now the most widely available route guidance system in the United States. A more recently released product, CARIN (car information and navigation) also integrates a GPS receiver, wheel sensors, an electronic compass, and navigable electronic maps. It incorporates many of the same functions as Navmate, including both audio and visual instructions, but CARIN adds cellular communications, which provides some additional benefits. Traffic and weather information will soon be available via cellular phone for integration into travel plans, and an SOS feature will be able to relay coordinates to authorities in case of an emergency. When the vehicle is stationary, the onboard monitor can also be used as a television. The system is currently available in the United States as an option on BMW 5 and 7 Series models.

### **Mobile Navigation Assistant System**

ADVANCE involves the use of a Mobile Navigation Assistant (MNA) system, a next-generation in-vehicle guidance system, developed by Motorola. Each vehicle is equipped with GPS capability, to

determine vehicle location; an RF communications system, to receive real-time traffic information and transmit vehicle location information; and a voice/graphics-based route guidance system. Route guidance is calculated by an in-vehicle computer. It receives information from the Traffic Information Center (TIC) through a radio frequency data modem. Because each vehicle is also acting as a probe, transmits traffic information (e.g., travel time data) from the vehicle to the TIC. The system receives information and sends data independent of the driver.

The Mobile Navigation Assistant contains the following:

- A 5.7-inch color (LCD) that provides the driver interface—with a touch screen and fixed buttons that allow the user to enter destination data, request specific functions, and access feature menus. A driver enters a destination by address, street intersections, or by choosing a point of interest resident in the on-board memory.
- Sensors including wheel speed, compass and differential global positioning satellite.
- A trunk-mounted computer that controls all vehicle systems.
- Single CD-ROM disk drive for mass data storage.

The MNA provides visual route guidance directions on the LCD display, and audio instructions through a speaker.

Some of the early findings of ADVANCE were:

- The density of probes need not be high for a probe-based system to be effective.
- The travel time reports provided by the probes were generally very accurate.
- Drivers can safely use in-vehicle systems while driving in traffic.
- The gap between potential and actual system performance must be closed before the confidence of prospective ATIS users can be won sufficiently to encourage them to buy the system and act consistently on ATIS guidance.

### **Radio Broadcast Data System**

A system known as Radio Broadcast Data System (RBDS) in the United States and as Radio Data System (RDS) in Europe is a way of sending information directly to wireless receivers. One of the initial uses for this technology is to broadcast accurate real-time traffic information directly to vehicles or smart information kiosks. Another use for RBDS is to broadcast differential correction information for use by global positioning system receivers.

Currently, traffic information is widely available through commercial broadcast radio; however, it does not always reach the traveler at a time or location where the information is useful. By using RBDS technology, timely traffic information is delivered directly to an instrumented vehicle or kiosk. Thus, relevant traffic information is always available to the traveler.

### The Phoenix Deployment Test

Through the Phoenix Deployment Test, real-time traffic information is gathered from the Arizona Department of Transportation, city of Tempe, Metro Networks, and Skyview Traffic Watch, Inc. Crusader Software (PC NT 3.51) developed by Daltek AB is used to encode the traffic information. Differential Corrections, Inc. broadcasts the encoded information from KSLX FM 100.7 radio station. Volvo systems (Dynaguide in-vehicle and PC-based units) overlay the traffic information onto Maricopa County Department of Transportation maps of the metropolitan Phoenix area. Retki software, developed by Liikkuva (running on notebook PCs), overlays the traffic information onto maps of the metropolitan Phoenix area.

Since 1994, General Motors has been offering the Duidestar in-vehicle units developed by Rockwell International and Zexel in its new Oldsmobiles, and will also offer the OnStar GPS-based guidance system from Delco/Hughes Electronics as an option on most of its 1997 Cadillacs at a cost of approximately \$1000 per unit. Numerous rent-a-car companies are also offering in-vehicle navigation units. Hertz and Avis are both offering the NeverLost system developed by Rockwell International and Zexel. National is offering the Siemens GPS-based units.

As discussed in the Regional Multimodal Traveler Information Systems section (Chapter 8), in-vehicle devices were used in the Atlanta Traveler Information Showcase for route guidance using real-time traffic information. In addition, The Hertz Rent-a-Car Corporation equipped 20 of its rental cars with the in-vehicle devices. Oldsmobile outfitted 30 vehicles of the fleet it provided to the Atlanta Committee for the Olympic Games with the devices. BMW provided the Showcase with five vehicles. The in-vehicle device were also placed in some private company fleets and Federal Highway Administration automobiles.

Currently, Acura, BMW, Oldsmobile, and Toyota offer navigation systems as optional equipment on some of their models. Sony, Rockwell, Clarion, Kenwood, and Amerigon systems can be installed in most older vehicles dating back to 1988. Hertz offers NeverLost (usually for an extra \$6 per day) on mid- and full-size rental cars as well as some Ford Explorers. The same system is available from Rockwell as the Pathmaster at car audio retailers. The Oldsmobile, BMW, and Toyota systems force the driver to scroll through a list of selections to choose the destination. Acura's system allows the driver to type a destination directly on the touch screen. BMW stores its navigation software on CDS in the vehicle's trunk. Acura and Oldsmobile use easily removable hard drives.

Two carmakers have combined cellular phone and GPS technology without on-board navigational software to create a different kind of system. Cadillac's OnStar and Ford's RESCU systems can pinpoint the vehicle's location within a few hundred feet and offer help when the motorist needs it. If, for example, a vehicle is in a collision severe enough to deploy the airbags, OnStar transmits a signal via satellite to a control center in Farmington Hills, Michigan, giving the location of the vehicle. An operator at the center will place a cellular call to the motorist of the vehicle to confirm the airbag deployment. If the motorist does not respond, the operator calls 911 in the area of the vehicle's location to dispatch police and an ambulance.

Besides summoning help if the airbags deploy, OnStar can remotely unlock the doors if the driver of a vehicle has locked the keys inside; can call the control center toll-free and the operator of the control center can sound the horn and flash the lights of a vehicle that the motorist cannot find; notifies the

control center if it detects that a vehicle has been stolen and the center then gives the location of the vehicle to the police.

The Ford RESCU's control center is in Dallas. The features of the RESCU system are similar to the OnStar with additional features: an icon appears, on the overhead console, for an ambulance and one for a tow truck. If the motorist pushes the icon button, the vehicle's location, and speed are sent to the center and the center will get appropriate help after verifying the call. RESCU, unlike OnStar, cannot remotely unlock a vehicle, flash lights to help motorist find the vehicle (Herb Shuldiner, March/April 1997).

RoadTrac in Roswell, Georgia introduced Ceres, a location-based vehicle security and monitoring product. Ceres uses GPS, cellular technology and a monitoring center to provide a variety of emergency and assistance responses to criminally threatening situations, accidents and breakdowns, medical emergencies, vehicle theft and mistakenly locked doors. In addition, it includes a remote key ring panic button which sends an alarm to the monitoring center and activates alarms on the car itself. It also offers a CD-ROM which allows the vehicle to be tracked on a personal computer (Inside ITS, January 13, 1997).

#### **SOURCES:**

- *The Art of In-Vehicle Navigation*, November 28, 1994, by Goug Wilsterman Intelligent Highway Systems, a Supplement to Engineering News Record
- <http://www.azfms.com/About/rbds.html>
- *Guiding Lights for Your Car*, March/April 1997, by Herb Shuldiner, Car and Travel

**[Return to Table of Contents](#)**