APPENDIX D

FUNCTIONAL ASSESSMENT OF TECHNOLOGIES

D

TABLE OF CONTENTS

APPENDIX -

D. FUNCTIONAL ASSESSMENT OF TECHNOLOGIES

APPENDIX D - FUNCTIONAL ASSESSMENT OF TECHNOLOGIES

A. Surveillance

Advanced Traffic Management Systems (ATMS) typically provide two different sub-systems for roadway surveillance: vehicle detection, and closed circuit television (CCTV). These two subsystems provide different functions, and operate together to provide the traffic operations center (TOC) with real-time status of traffic conditions. The vehicle detection sub-system electronically monitors the flow of traffic on the roadways, and transmits this information in "real-time" to the TOC for analysis and status displays. The operators utilize the results of the analysis and the status information to make decisions regarding management of the traffic. The CCTV sub-system provides the operators with visual means for verification of the conditions reported by the vehicle detection sub-system. The CCTV images also provide the operator with an independent evaluation of traffic conditions.

Each of these two sub-systems can be deployed and utilized jointly as well as separately. However, the complementary interaction of the two sub-systems improves the overall system operation in a manner that neither system can provide alone. The vehicle detection system, since it is automated and can function with minimal human intervention, provides continuous surveillance and up-to-the-minute data. The CCTV system allows the human observer to view and interpret an incident, or other traffic conditions, and determine an appropriate response. As more progress is made in the technologies of image processing, artificial intelligence, and expert systems, it is inevitable that computer systems will augment the capabilities of the human observer.

1. Vehicle Detection

Vehicle detection technologies form the foundation of the surveillance sub-system used for automated incident detection and traffic management. Surveillance information provided by vehicle detection enables collection of a range of traffic data including speed, volume, density, travel time, and in some cases, vehicle position. Control strategies, incident management procedures, and motorist information displays are selected based upon the data collected by the vehicle detection system. The collected data is used in real-time for making traffic management decisions and stored to provide a historical data-base of traffic conditions. Surveillance can also be used to obtain information on vehicle classification, length, speed, acceleration characteristics, and hazardous materials.

Operational environment and maintenance requirements are two of the most critical factors in determining the types of detectors for the system. Systems that involve cutting existing road surfaces and pavement (such as induction loops) can create installation and maintenance problems or compromise the structural integrity of the roadway, especially on bridges and other structures. Technologies that do not require these modifications are termed non-intrusive installations, and minimize the traffic diversion and control problems.

The choice of detectors for an automated system is dependent on the data requirements. To meet the needs of the recommended system, real-time data to ascertain vehicle speed, counts, lane occupancy, classification, and changes in motion and position, will be required for automated incident detection. This real-time data should be stored for historical as well as planning use.

There are two separate approaches to vehicle detection; those that are passive and involve no electronics in the vehicle, and those that cooperatively utilize electronics in the vehicle and alongside the roadway. As with all areas of electronic technologies, changes occur regularly providing new solutions to existing problems, but conversely requiring that systems be flexible enough to accommodate change on a regular basis.

Various technologies applicable to this project are discussed below. There are numerous other technologies that have been experimented with and tested by various DOTS and the FHWA. In particular, the current "Detection Technology for M-IS" project sponsored by FHWA is evaluating a wide range of equipment under laboratory and field conditions. Although many of these technologies show promise, they have not progressed to reliable field operation. In order to limit system complexity, and resultant operations and maintenance costs, minimizing the number of different technologies is preferred.

Passive Vehicle Detection

Technologies that do not require any devices in the vehicle are the basis for most current vehicle detection systems. Passive approaches allow all vehicles in the vicinity of the sensor to be detected and monitored, but provide less information than will ultimately be available in the future.

Induction Loops: The most commonly used vehicle detection technology is the induction loop. This technique is extensively used for arterial controls and has a long history of successful field deployment. The advantages of induction loops are their well-known performance characteristics, maturity, application flexibility, and multiple vendor availability. Over the years the manufacturers have enhanced and refined their equipment, providing numerous options and alternatives to meet a wide range of application needs. Pairs of loops can be used to measure speed and vehicle length for classification purposes. Some vendors have announced products that measure speed with a single loop, but field experience is limited. Some disadvantages of induction loops are the result of the need to embed the loops in the pavement surface, and the problems associated with pavement deterioration and freeze-thaw damage. Further difficulties include damaging the loop conductors during resurfacing operations or construction, and the reduced effectiveness of loops when in close proximity to reinforcing steel.

Recent improvements have been made in inductive loop technology. Loop detectors have been primarily utilized to provide a digital output that is representative of vehicle presence above the induction loop in the pavement. In this regard, a sophisticated computer system is unable to gain access to any information contained in the magnetic or inductive signal collected by the detector amplifier. New products are available with on-board microprocessors that are able to monitor the "signature" of the detected field. Use of this data allows accurate speed measurement and provides some capability for classification. Serial data ports with RS-232 communication, allow systems to access a detector amplifier internal database, to perform remote sensitivity adjustments and compensate for weather conditions.

Another development is the manufacture of pre-formed loops, which are available from a number of suppliers. This type of loop is pre-assembled, with the wires encased in a filled conduit. This assembly is embedded in the pavement, typically several inches below the surface, during the construction of the roadway. This technique offers improved reliability and life expectancy.

A similar approach, that of embedding the loop in the pavement, is being utilized in some areas as part of roadway reconstruction projects. After the milling operation that is used to remove old pavement, the induction loop is saw-cut into the milled surface. After the new pavement is applied, the loop is buried several inches below the road surface, where it is less subject to damage from traffic, construction or weather.

While induction loop detectors are often maligned because of the problems noted previously, they are currently the primary source of vehicle detection in most systems around the country. Studies in Los Angeles were performed by video taping the traffic stream, time-stamping and manually counting the vehicles on the video tape. Results show that the accuracy of induction loop data with respect to vehicle counts is +0.6%.

Magnetometers: Magnetometers, and the related micro-loop technology, are often suggested for deployment on bridges and other areas where loop installation in the existing pavement area could affect structural integrity. Magnetometers have had spotty operational success, and other technologies have often been considered for these particular needs. However, the use of new digital processing technology has the potential to significantly improve the performance of magnetic detectors. A re-evaluation of their role will be appropriate after sufficient field experience is gathered. Preliminary results from the IVHS Detection Technology project show that magnetometers have an accuracy in the +/- 5% range.

Axle Counters: The FHWA requirement for 13 bin vehicle classification on certain roadways generates the need to count axles. The most commonly used technology uses a bending beam piezoelectric strip embedded in the roadway surface. These devices, working in conjunction with inductive loops, measure the vehicle length and speed, and count the axles. The vehicle length, combined with axle count, are used to classify the vehicle.

Radar: Radar detectors operate by emitting a signal in the microwave portion of the electromagnetic spectrum, and analyze the returned signal. These detectors are in limited use in incident detection and freeway management projects. Continuous wave (CW) radar detection operates on the Doppler effect (measuring frequency shifts between the transmitted and received beam caused by vehicle motion), and thus directly measures vehicle speed. Vehicle counts can be determined by accumulating each vehicle detected, but this approach cannot readily obtain lane occupancy and vehicle lengths. Similarly, detection of stopped vehicles, or very slowly moving vehicles, is difficult.

Another type of CW radar detector transmits a signal that is swept over a range of frequencies. This technique allows measurement of range from antenna to vehicle, and is thus able to function as a presence detector. The sweep frequency however functions much as sample rate to quantify presence.

Pulsed radar operates by transmitting a burst of microwave energy, and interpreting the "echoes" reflected from vehicles in its "field of view". Because of the complications involved in processing multiple reflection, pulsed radar units utilized for traffic detection limit their field of view to a portion of the lane, such that a single vehicle is present in their detection zone. This technique permits the determination of the distance to the nearest reflection, and by monitoring this reflection over time, the position and resultant speed of the vehicle can be determined. This type of radar can be used to sense stopped vehicles, but has the limitation of a sample rate that must be frequent enough to provide accurate presence calculations to determine other traffic parameters.

Continuous wave radar detectors of the Doppler and swept frequency types require one antenna per lane, mounted on a structure or a sign bridge over the lane. The same limitation applies to pulsed radar units. The IVHS Detection Technology project early results show that these radar detectors have accuracy's that range from +/- 5% to +/- 6%

A new technology ***microwave detector*** has recently become available. When mounted at the side of the roadway, this device is able scan up to twelve lanes. Since side mounting facilities are often available, or can be readily installed, the device is more cost effective. The device can also detect vehicle presence, and is thus able to determine occupancy and existence of stopped vehicles. However, it does not measure speed directly, relying upon "single loop" speed estimation techniques based upon average vehicle length. The accuracy of this device, as stated in the early results from the Detection Technology project, is in the +/- 5% range for volumes. Test results indicate missed and duplicated counts across multiple roadway lanes upon the passage of large vehicles.

The advantages of radar and microwave devices include the ease of use, requiring no cutting of pavement and disruption of traffic flow for installation or maintenance (if mounted on a structure or sign bridge where overhead access is possible). For the Doppler units, direct speed measurement is a significant benefit. If traffic lanes are shifted, radar antennas can be easily re-aimed. The disadvantages of radar are: the overhead mounting requirement, limited field operational experience for many of the new units, a small number of vendors in the market, and difficulties of accurately sensing lane occupancy and slow moving or stopped vehicles with Doppler units.

Radar detectors can be configured with two types of interfaces: RS-232 serial data and two pulse-type contacts. The serial output provides data (volume, speed, etc.) in an ASCII text string. Modifications to this format to incorporate an error checking communications in a standardized protocol would allow a multi-lane unit to be installed without a local field microcomputer. The dual pulse-type contact closures provide for emulation of a loop-pair speed trap. The first contact closure occurs when the vehicle enters the detection zone, and the second contact closure is timed relative to the first closure by the detector to provide the correct travel time based upon a calibrated “loop spacing”.

Infrared: Infrared detectors monitor electromagnetic energy in the band above the visible spectrum. Both active and passive devices are marketed that utilize infrared detection.

Active infrared devices illuminate the detection zone with infrared energy supplied by either light emitting diodes (LEDs), or lasers. Lasers can provide a higher level of output energy. A portion of the energy reflected back from the vehicle is detected and processed. The detector consists of optical elements to focus the returned signal onto a matrix of infrared sensors. The two-way travel time of the infrared pulse from the source to the sensors is used to measure the distance to the vehicle. This strategy is similar to that used in a pulsed radar detector. Processing of the data provides vehicle counts, occupancy, presence, speed and classification information, Because infrared energy is attenuated and scattered by rain snow, fog and mist in the air, active infrared detectors are vulnerable to these atmospheric conditions. In addition, other obscurants in the air, such as smoke and dust, can reduce the effectiveness of the detector.

Passive infrared devices do not emit any energy themselves, but utilize the characteristic that all objects emit heat (infrared radiation) as a function of their surface temperature. The amount of infrared energy is also a function of the emissivity of the object itself By detecting difference between the temperature/emissivity of vehicles and the roadway surface, a passive infrared detector can determine the presence and passage of vehicles. The infrared energy is focused through an optical system onto the infrared sensors. The resultant signal is processed to provide presence, vehicle counts and occupancy. As noted above, infrared energy is obscured by atmospheric effects. Because passive infrared detectors are dependent upon the sun and other infrared sources for their input energy, diurnal changes, cloud cover, glint from bright objects reflecting sunlight, etc. can create confusing and unwanted signals.

By increasing the number of sensors in a passive infrared detector, an “image” of the scene of interest can be generated. This increase in detail allows additional information from the scene to discerned and analyzed. As the number of individual sensors becomes large enough, the boundary between an infrared detector and an infrared sensitive CCTV camera becomes blurred. For practical applications, an infrared imaging system has essentially all the same characteristics of Video Image Detection systems discussed below.

Sonic: There are several techniques that have been explored utilizing sound. Some devices operate as sonar devices, sending out sound waves and analyzing the returned echoes from the vehicles - much like the radar systems. The early results from the ultra-sonic unit included in the IVHS Detector Technology test show an accuracy in the +/- 2% range. Other sonic detectors passively "listen" to the noise generated by the vehicles, and analyze this noise energy to detect individual vehicles and resultant location and speeds. These devices have not yet been extensively used, and thus field experience is limited. However the technology has been applied for submarine noise signature detection by the military and could become a valid tool for classification of vehicles.

Video Image Detection: Video Image Detection (VID) systems (sometimes referred to as machine vision systems) are comprised of fixed orientation CCTV cameras strategically located to provide views of specific areas or long sections of roadway, coupled with a computer that analyzes the video image in real time (30 times per second). This technology has been developed for various industrial, manufacturing, military and aerospace applications. It has been applied to traffic management in recent years, with growing success. Early systems were troubled by harsh environments, adverse and changeable lighting conditions, shadows, differing vehicle shapes, and sometimes difficult operating conditions. These difficulties have, for the most part, been solved by extensive field testing, actual deployment, more powerful computers, and increasingly sophisticated software.

Two fundamentally different strategies are used to analyze the video images: fixed analysis zones that detect vehicles moving through them, and vehicle identification and subsequent tracking. A third strategy, involving reading license plates "on-the-fly" may be appropriate for toll violations and related applications, but is not directly applicable to this project. The technique utilizing fixed analysis zones, analogous to a "loop" in the video image, is the most stable and best tested approach. Equipment based on this approach can provide vehicle counts, lane occupancy, speeds, and lengths. Software in the VID processor collects the standard parametric information (volume, occupancy, and speed) and can also provide some analysis and processing of this data, including statistics accumulation, data smoothing, and level of service calculations.

A key benefit of a VID system is its ability to monitor large areas of roadway from a single equipment location. Because the CCTV camera can be oriented to monitor a section of roadway (up to l/4 mile in length), and the entire image can be analyzed, significantly more roadway and numbers of vehicles can be monitored. The most promising usage of a VID system is detection of stopped or stalled vehicles (either in a travel lane, or on the shoulder), providing direct detection of an incident. The monitoring of wide areas of roadway, coupled with individual vehicle detection, will provide significantly more information than existing point source (such as induction loop or radar) technologies.

While the promise of VID systems is significant, it is still a young technology that will evolve and grow for many years. There are operational problems under adverse lighting, transitions between daylight and darkness and storm conditions that will require more refinement. Camera placement must be carefully considered, as shadows from objects outside the detection area may affect performance. The early results from the Detector Technology project report show accuracy's ranging from +/- 0.3% to +/- 2.3%, with accuracy decreasing under dark or adverse weather conditions.

Passive Vehicle Detector Cost Comparisons

Two different categories of passive vehicle detectors are discussed above: those that are embedded in the road surface, such as induction loops, and those that are mounted overhead, such as a radar detector or a video image detector.

As discussed, embedding detectors in the roadway requires that the road surface be cut or drilled, and subjects the detector to failure due to pavement deterioration, etc. This can create ongoing maintenance problems, or poor detector reliability. As noted, newer construction techniques which embed the detector several inches below the pavement surface are being used to solve some of these problems.

Detectors that are mounted above each lane, such as most radar detectors and ultrasonic detectors, require some form of support structure. A claimed advantage of this installation location is minimal traffic disruption during installation and maintenance. Mounting on an existing overcrossing is an option, but can create aesthetic concerns and often results in limited accessibility requiring that a traffic lane be closed to service the unit. The use of signal head mast arms is another possibility, but has the drawback of motion under high wind loading and the need to block traffic for installation and servicing. Sign bridges are a third possibility, and where they already exist are excellent choices, especially if they include a cat-walk so that the units can be installed and serviced without shutting down traffic However, the installation of new sign bridges for the mounting of detectors is an expensive alternative.

In general terms, many of the overhead detectors cost between $750 and $1000 per unit that monitors a single lane. Poles and mast arms cost about $200 per foot (with foundation and installation), resulting in a cost of roughly $2400 for a 12 lane. This is about 2.5 times the cost of the detector. Sign bridges roughly cost $500 per foot (with foundation and installation), or $6000 for a 12 foot lane. This is about 6 times the cost of the detector. This needs to be compared to the installed cost of induction loops of about $1000.

Thus, overhead mounted detectors that must be positioned over each lane can be significantly more expensive than induction loops, when the cost of a mast arm and pole, or sign bridge must be included. Under those situations where an existing structure or sign bridge is available, they can be cost effective - but may still require traffic disruptions for installation and servicing.

Another category of overhead devices - side fired radar and video image detectors (VIDs) - can be mounted off the side of the road or on a pole in the median. This reduces the cost of mounting to roughly $5000, and does not require stopping traffic for access to the unit.

These devices also have the advantage of being able to monitor several lanes from a single unit, thus spreading the cost of the unit and the mounting pole across several lanes. A disadvantage of side mounting or an oblique camera view is the blockage of line of sight by larger vehicles (trucks) of smaller cars. This results in missed counts. With VIDs, the ability to discriminate between two closely following vehicles is a function of mounting height and angle of view. Increased height improves the discrimination ability, but results in a more costly pole and foundation. Another problem noted with VIDs is motion of the mounting pole under wind loading, or twisting of the pole due to differential solar heating. These conditions result in the camera field of view changing and "moving" the fixed analysis zones to another portion of the image.

For comparison purposes. a six lane cross section of freeway has been utilized as shown on Table D-1. Five different equipment configurations have been evaluated:

- Induction Loops, with lead in wires saw-cut into pavement surface and processor cabinet on one shoulder;

- Side Fired Radar, with unit mounted on a pole located on one shoulder adjacent to the processor cabinet;

- Video Image Detector, with two cameras mounted on a pole in the median and the processor cabinet on one shoulder;

- Overhead Mounted Sensors on Mast Arm with the pole in the median and the processor cabinet on one shoulder; and

- Overhead Mounted Sensors on Sign Bridge with processor cabinet on one shoulder.

For all configurations, it is assumed that power and communications conduits are available at the location of the processor cabinet. With the exception of the video image detector, a Model 170 processor and cabinet is included. Conduit, cable, installation and testing costs are included for all cases. For the two configurations with median located poles (VID and Overhead Sensors on Mast Arm), costs for jacking conduit under three lanes are included.

TABLE D-1

PASSIVE VEHICLE DETECTOR COST COMPARISON

CONFIGURATION	PER LANE	SIX LANES
Induction Loops	***$3,400***	***$20,300***
Side Fired Radar	***$3,725***	***$22,350***
Video Image Detector	***$10,100***	***$60,600***
Overhead Mounted Sensors on Mast Arm	***$6,250***	***$37,500***
Overhead Mounted on Sign Bridge	***$13,250***	***$79,500***

Maintenance costs are usually calculated as 10% per year of the equipment costs. Induction loops may be higher if local experience shows that typical loop life is short.

Active Vehicle Detection

Technologies that place electronics in the vehicle that interacts with the roadside infrastructure, and other vehicles in the immediate vicinity, is the direction of progress for automated guidance and highway systems. It will be at least two decades before these technologies become widespread, but devices in this category are being used for specific applications around the country.

Automatic Vehicle Identification The recent conversion of many toll facilities to electronic toll tags for electronic toll collection (ETC), also referred to as automatic vehicle identification (AVI), creates a potential for vehicle detection and monitoring. By monitoring the movement of individual vehicles past various AVI antennas, the vehicles become active probes and link travel times can be determined. This technology is successful in areas where AVI tags are in use for toll roads, but is of limited applicability elsewhere. The installation of AVI on portions of the New York State Thruway, and the Thruway's membership in the EZPass Interagency Group, will eventually result in a population of AVI equipped vehicles in the Rochester area.

Another use of AVI technology is its use on transit vehicles to determine their location. The use of induction loops as the reading antenna has been successfully deployed in some areas. This usage of AVI has found a receptive audience as a method for more accurate tracking of bus fleets for control and dispatch.

Global Positioning Systems (GPS): GPS equipment is being used by various emergency (police, etc.) and fleet (trucking) organizations to permit continuous tracking of vehicle locations. The costs per vehicle are still too high for widespread usage by the general public, but the technique is very beneficial for those cases where it can be justified. Accuracy's range from a few hundred feet, to a few feet, depending upon the capabilities of the GPS receiver. The more accurate units are proportionally more expensive. GPS receivers as accessories for PCs are now available at prices of less than $1000. As sales volumes increase, prices will continue to come down and additional hardware and software features will be added.

GPS receivers are an important component of vehicle navigation systems currently being tested. It is included as a component of the in-vehicle navigation systems and vehicle emergency notification (Mayday) systems being considered as part of the National IVHS Architecture being developed by the USDOT. Vehicle location using this technology, coupled with a data channel linking a public service vehicle (police, fire, transit, etc.) to the TOC is being evaluated as a component of incident response systems elsewhere in the US. The ability to locate emergency response vehicles in real time on a status map, is a very useful tool in managing and coordinating incident response over a wide area. After some initial operational experience is gained from systems currently in development, the effectiveness and costs can be evaluated for possible use.

Automatic Vehicle Location: A variation on the GPS strategy is the use of fixed location beacons that can be monitored by a vehicle, such as a bus. Through the use of an on-board computer, monitoring of the vehicle's movement with an electronic odometer, and known information about a route to be followed, the location of the vehicle can be estimated. The location beacon allows the strategy to be refined by providing "check-points" that permit the on-board computer to update and correct its estimates of location.

The periodic transmission of vehicle location to a central computer allows a central dispatcher to track the vehicle. This tracking can be matched to a bus schedule, for example, and alert the driver and the dispatcher if the bus is ahead of or behind schedule. This automated vehicle monitoring can be input to the traffic management system. It would provide active probes in the vehicle stream, similar to the AVI system discussed above. The use of buses as probes must take into account the start/stop nature of transit vehicles when estimating the flow of traffic. The integration of this tracking with voice communications to the bus driver is a very useful tool in locating incidents, and determining their nature and severity.

Detector Comparison Matrix

Table D-2 illustrates the major features of the most common types of vehicle detectors. The matrix is a tabulation of the major classifications of detector types. It includes the primary parameter that is most directly measured by the detector and the preferred mounting.

TABLE D-2

DETECTOR COMPARISONS

DETECTOR TYPE	PRIMARY DATA	MOUNTING LOCATION	COMMENTS
LOOP	PRESENCE	ROADWAY PER LANE	Roadway cut installation life apx. 3 yrs.
PIEZOELECTRIC	AXLE COUNT WEIGHT	ROADWAY PER LANE	Installation involves roadway cut.
RADAR (CW)	SPEED	OVERHEAD PER LANE	Poor results at low speeds.
RADAR (MULTI-ZONE)	PRESENCE	OVERHEAD, SIDE MULTILANE	Some tests show difficult calibration.
PASSIVE IR (NON-IMAGE)	PRESENCE	OVERHEAD PER LANE	Few installations.
PASSIVE IR (IMAGE)	PRESENCE	OVERHEAD, SIDE MULTILANE	Few installations.
ACTIVE IR (NON-IMAGE)	PRESENCE	OVERHEAD PER LANE	Few installations.
ACOUSTIC (PASSIVE)	PRESENCE	OVERHEAD PER LANE	Some tests show reliable operation. Few installations.
ULTRASONIC (PULSED)	PRESENCE	OVERHEAD PER LANE	Poor sample rate for high speed flow statistics.
ULTRASONIC (CW)	PRESENCE	OVERHEAD PER LANE	Poor results at low speeds.
MAGNETOMETER OR MICROLOOP	PRESENCE	ROADWAY PER LANE	Manufacturer claims good results on bridge decks.
VIDEO IMAGE	TRACKING	OVERHEAD, SIDE	40 ft mounting height suggested.
AVI	TRAVEL TIME LOCATION	OVERHEAD, SIDE LIMITED RANGE	As electronic toll use increases, population of users should grow.

2. Closed Circuit Television

Closed circuit television (CCTV) provides the eyes for the operator at the traffic operations center, and has proven to be one of the most valuable elements of an ATMS. Operational experience shows that constant monitoring of CCTV images by operators is not effective, as the operator soon becomes “numbed” by the constant repetition of vehicles moving across the screen. The primary role of CCTV is to verify a reported incident or other traffic condition, to evaluate its severity, and determine the appropriate response vehicles and personnel to dispatch to the incident scene.

In addition to its primary role in incident verification and response coordination, CCTV can also be used for purposes in the Rochester ATMS. These include:

- Monitoring the operation of critical signalized intersections that are in the vicinity of the CCTV camera. This allows evaluation of signal timing and the related functions of the controller. One agency has reported the installation of a spare optical fiber to each intersection so that they can install a CCTV camera on an as-needed basis during trouble shooting and problem isolation. This saves them many trips to the site when they are trying to correct intermittent failures.

- Utilizing the CCTV camera to monitor adjacent parallel streets to a freeway to determine current operating conditions. This allows verification that the arterial streets have adequate vehicle capacity to handle added traffic prior to implementing a freeway diversion plan. Monitoring of the operation of the streets during the diversion to insure successful operation is also available.

- Monitoring motorist response and traffic movements on the mainline, entrance and exit ramps, and HOV lanes. This is utilized to verify compliance with ramp metering or HOV restrictions, or observance and response to messages posted on a VMS or transmitted on an HAR.

CCTV cameras, lenses and typical mounting heights (40 feet above the roadway surface) allow monitoring of roughly one-half mile each direction from a camera location. This is of-course restricted by topography, roadway geometry and vegetation. Some installations have mounted CCTV cameras on high-mast poles or towers more than 100 feet above the road. This added height provides larger areas of coverage, if topography and vegetation are favorable.

Specific selection of camera locations is controlled by the desire to monitor high-incident locations and other areas of interest. Ability to view parallel surface streets and ramps should also be considered in site determination. The constraints imposed by access, available locations for cabinets and pole foundations, and communications often limit the optimum selections. Each prospective site must be investigated to establish the camera range and field-of-view for the mounting height and lens combination selected.

The biggest problem to overcome with CCTV is the transmission of the image from the camera location to the control center. Direct video requires a communications channel that is equivalent to more than 1500 voice grade audio channels. Thus, most efforts in optimizing CCTV systems are directed toward reducing the bandwidth of the CCTV communications channel, These efforts range from not updating the image in real-time (every l/30 sec) to digitally compressing the image, through analyzing the image and transmitting only the moving elements of the image.

The standard for CCTV pictures is a "broadcast" quality, full-motion, real-time image. At present, this is usually implemented by use of a fiber optic communications system, with a separate full bandwidth fiber allocated for each CCTV camera. With tremendous bandwidth available on a fiber optic system, this direct approach is often the least costly and provides the best performance. When this direct approach is not cost effective, alternative solutions must be utilized.

Camera Type

Color images provide the greatest amount of visual information, and are the preferred choice of most traffic operations centers. However, color CCTV cameras rapidly lose their sensitivity under nighttime, or other dim, lighting conditions. Black-and-white cameras, on the other hand, are available that will produce usable images even when it is too dark for a person to see. A black-and-white camera is able to produce a usable image with 1/10 or less) the light level required for a color camera. Some vendors have solved this dilemma by packaging both a color camera and a black-and-white camera in the same housing. This of course increases the price of the assembly, but the added cost may be acceptable in some locations. Actual field testing should be performed, or verification of performance of cameras at existing traffic operations centers, before committing to a specific equipment selection. The typical cost of a color camera, with field controller and cabinet, pan/tilt unit, housing and mount, with installation and testing is roughly $20,000.

Psn/Tilt/Zoom/Focus Control

The CCTV camera in the field must be moveable (left and right; and up and down) in order to permit it to monitor the greatest possible area. Similarly, a zoom lens to allow viewing of vehicles at varying distances and associated focus control, is required. These functions must be controllable by all operators who have access to the CCTV images. This functionality is implemented by placing a microcomputer at each CCTV location that receives commands from the traffic operator and turns on and off the appropriate motor in the pan/tilt unit or the motorized lens.

Each CCTV system vendor has its own proprietary system for this type of control. As systems grow and expand over time, control compatibility must be maintained so that the operator is not faced with several different camera control systems. The needs for the control system, both initial and long-term must be addressed during the system architecture design, considering the growth requirements and future needs.

Digital vs. Analog Transmission

The technology used to date for most long haul, "broadcast quality" CCTV systems has been analog transmission. Within the past five years, significant progress has been made in the development of cost-effective digital transmission equipment. Once video is converted to the appropriate digital format, it can be transmitted long distances over a fiber optic link using a digital protocol such as a Synchronous Optical Network (SONET) communications system with no further conversion and without degradation of image quality. Additionally, digital video switches are smaller, and lower priced than analog switches.

Another benefit to digital video is the ability to compress the video image, and thus utilize lower bandwidth on a less expensive data communications channel, which may be used to transport the video to another facility. Typical compression ratios are 40: 1. The cost of compression/ decompression codec equipment is currently about $20,000 per unit, but new products are being discussed which may bring the price down into the $5,000 range. Given normal price/performance curves in the digital electronics industry, this price drop will probably require about 3 years. However, if the price/performance ratio of digital systems does not progress as rapidly as desired, an analog system will provide fully satisfactory results.

Fiber vs. Coaxial Transmission

The use of fiber optics for transmission of video has almost completely replaced the use of coaxial cable, except for very short runs of less than 500 ft. Disadvantages of coaxial cable include requirements for amplification of the transmitted signal every few thousand feet to compensate for signal attenuation, and the susceptibility of the cable to induced noise. Fiber optic transceivers are now available with ranges up to five miles for multi-mode fiber, and over 20 miles for single-mode fiber. These transceivers range from less than $300 for short range units to over $2000 for long range devices.

Geographically Distributed Control

An effective and needed strategy in modem incident/traffic management systems is distributing video images to the multiple locations and agencies that can utilize them. This provides for joint, coordinated response to an incident. in addition to the video images, camera selection and pan/tilt/zoom control must also be distributed. Geographic distribution of these control functions must be considered in the basic design of the CCTV system, since adding these capabilities to a simpler system is often difficult and costly.

Video Switch

A key component of the CCTV system is the video switch that allows any CCTV camera to be viewed on any monitor, at any location that has access to the CCTV system. A variety of switch architectures are available, from fully centralized to fully distributed. Each has its own advantages and disadvantages, and associated costs. Most CCTV systems have more cameras than monitors, with typical ratios being in the 3: 1 to 10: 1 range.

The cost of analog video switches is a function of the number of switching points, which is the product of the number of camera inputs and monitor outputs. Thus, prices can increase exponentially as the size of the switch grows. For a relatively small switch (30 camera inputs and 10 monitor outputs) the installed cost is about $20,000. Doubling the size of the switch to 60 camera inputs and 20 monitor outputs results in the cost increasing to about $75,000.

Newer digital techniques, similar in concept to a LAN, are being utilized to transmit and switch video images. With these techniques, the video image is digitized and divided into small segments. These segments (or packets) are then distributed on a very high speed transmission system, and those users who need to view a particular image copy the packets for that image and reassemble the image for viewing. This strategy is commonly used in the telephone industry for switching voice conversations. Since switches of this nature do not increase exponentially in size, they have the potential for being less costly than analog switches. However, because of the high bandwidth and transmission speeds required, these devices are still more expensive than moderate sized analog switches. With the typical decline in costs for all digital based systems, digital switching of video images will rapidly become a cost effective alternative.

In all cases, cost of video monitors, interconnection to the video transmission system and monitors, operator controls and system integration is in addition to the cost of the basic switch.

Large Screen Video

A large video screen (Often 3’ x 4’, or larger) is frequently included in traffic operations centers. The ability to project either an enlarged video image, or an enlarged computer generated graphic can be useful for decision support during incident response or for public relations during tours or demonstrations. Operators in TOCs with large screens report that they seldom use these enlarged images during normal operations.

Two fundamental technologies are available: video projection, and video wall. Video projection utilizes either a CRT or an LCD system to optically enlarge the image and display it on a screen (using either front or rear projection). Care must be exercised with the room lighting as the projected image is easily washed out by available light. A video wall provides a large display area that overcomes this problem. The video wall combines a number of moderate sized (21 inch typical) video monitors into an array. This array is often four monitors high and four monitors across. Electronic circuitry divides the original image into smaller parts (say sixteen for a 4 x 4 array) and displays each sub-image on a separate CRT. Current cost for large screen projectors is in the $35,000 range, while video walls are often above $50,000.

3. Electronic Toll and Traffic Management

Electronic Toll and Traffic Management (ETTM) systems encompass Automatic Vehicle Identification (AVI) and Electronic Toll Collection (ETC) with communication between vehicle and roadside. Transponders carried by vehicles participating in the program can be used to track travel times of vehicles on the roadway. The information obtained in this manner can be used to improve detection of incidents that create significant impacts on the level of service provided by the roadway system.

The use of Electronic Toll Collection (ETC) is expected to become widespread in the New York, New Jersey, Pennsylvania area in the next few years. Electronic tags, slightly larger than credit cards are placed on the inside of a vehicle windshield. The tags communicate on high frequency radio links with equipment at a toll plaza as the vehicle passes through the toll lanes. Violators using the lanes without a valid tag cause an image of the vehicle license plate to be processed by an agency that will send a bill or summons to the registered owner.

This "high-tech" scenario is presently in use at the Tappan Zee Bridge and various other barrier tolls on the New York Thruway. Several other agencies have also begun ETC operations.

Inter-Agency Group

In order to standardize use of ETC in the New York, New Jersey, Pennsylvania tri-state area, the EZ Pass Inter-Agency Group (IAG) was formed. The IAG consists of

- MTA Bridges and Tunnels (formerly TBTA)
- New Jersey Highway Authority
- New Jersey Turnpike Authority
- New York State Thruway Authority
- Pennsylvania Turnpike Authority
- Port Authority of New York and New Jersey
- South Jersey Transportation Authority

The systems installed by members of the IAG will use compatible electronic tags and will use a unified customer service center base. The tag technology that has been selected by the IAG is MARK IV.

Field trials

Experimentation by the New York State Thruway Authority to this point has been with AMTech tags. Initial facilities near the New York City area have been established at the Tappan Zee Bridge and Spring Valley barrier tolls. The tags are also in use at two barrier toll plazas in the Buffalo area. With IAG selection of MARK IV technology, a complete change over of toll-tags for each participant and replacement of electronic equipment for each toll lane for electronic toll collection to correspond to the new EZPass/MARK IV standard at will performed in mid-1995.

Coordination of user accounts with multiple agency clearinghouses is still not fully resolved by the IAG. A driver who intends to use several toll facilities may still need to establish a separate account with each agency before the EZPass tag is valid for use on the multiple facilities involved.

Vehicle Probes

ETC provides the opportunity for vehicle tracking that was formerly not possible. Each EZPass tag has a uniquely identified electronic serial number and can be read at highway speeds. The use of equipment to read the tags at opposite ends of a highway link allows the system track the passage of individual vehicles and thus provide a direct measurement of link travel time. Each equipped vehicle provides the system with an origin location and time and a destination location and time for each equipped link and hence becomes a probe vehicle without disturbing the traffic flow.

The Port Authority of New York and New Jersey has recently installed an Automatic Vehicle Identification system at Kennedy Airport to track airport buses. The system uses tags which were intended to be the same as EZPass tags, however the system was based on earlier thoughts of the IAG when they seemed to favor another technology provided by AMTech. As the airport system currently only tracks buses as a stand-alone system, there are no plans to convert the technology for uniformity at this time.

Installation of the antenna to communicate with the in-vehicle transponder must usually be done so that the communications range can be kept below a distance of about 30 feet. An overhead antenna may be able to cover three or four lanes simultaneously. The question of interference from multiple tags responding simultaneously has been considered by the various manufacturers. A solution proposed by MARK IV uses an overhead antenna array to communicate across the multiple lanes. In this configuration, one reader is connected to the multiplexed array of radio frequency transceivers which are in turn connected to separate antennas for each lane. For the purpose of collecting travel time data, a lower cost “compact reader” interfaced to only one antenna could be used to scan multiple lanes. Some vehicles will be missed when in the shadow of another vehicle or simultaneous responses cause the return data to be garbled, and some vehicles will be missed if they are in a position that is beyond the range of the equipment. Data collected from a few such stations should however provide enough successful matches of transponder reads to calculate a good estimate of travel time without performing a complete census of the EZPass users on the roadway.

Alternative Probe Technologies

Investigations are being performed that utilize cellular telephone serial number tracking with direction fmding capability at some cell towers to determine average vehicle speeds.

Legislation concerned with privacy may have an effect on this type of device, but products that do not monitor conversations and do not use actual telephone number identification are still in development. A demonstration project is currently in construction (with funds from the FHWA) for the Virginia Department of Transportation (VDOT). This project employs a system that will compute vehicle speed vectors from a combination of direction finding and map matching. The system should be in operation by the end of 1994.

B. Traveler Interface

1. Variable Message Signs

Variable Message Signs (VMS), both fixed and portable, are widely used to provide motorist information during an incident. The ability to quickly alert motorists of a problem ahead, and provide for diversion to an alternate route, is a successful strategy for minimizing the impact of an incident.

A VMS consists of a matrix of dots, each of which can be individually controlled. The minimum group of dots for a single character is five dots horizontally and seven dots vertically. Larger "character cells" are often implemented for improved character resolution, the use of lower case letters, and "double stroke" characters. Since individual characters on a VMS are composed of discrete dots, the "sharpness" of a character is controlled by the number of dots per character. The tradeoff is cost, with cost of the sign being proportional to the number of dots on its surface. The human eye fuses together the adjacent dots in the character pattern, and recognizes the character as a whole. In general, the legibility of 5 inch by 7 inch character cell VMSs is very acceptable, especially if only upper case letters are used, which is typical for roadway applications. When lower case is required, or other effects are needed, larger character cells, and proportionally more expensive signs, are necessary.

If the VMS is intended for text messages only, adjacent "character cells" can be separated by a blank space to minimize the cost of the sign. An alternative approach is the "continuous matrix" sign, in which the separating blank space is deleted, resulting in all locations on the surface of the sign being controllable. This permits moving text, "exploding" and "collapsing" images, roller blind, horizontal shutter, and other types of special effects to be implemented. These special effects are more commonly used in commercial displays than in roadway applications. Use of a proportional font for improved readability or graphics is a common use of continuous matrix signs on a roadway.

Various display philosophies are in use by different agencies. Some feel that a VMS should only be used when necessary to display instructions or information about roadway conditions, feeling that if routine messages are displayed, the driver's awareness of the sign becomes numbed. Other agencies display a routine or safety message on the signs to confirm operability, while some agencies use their signs to advertise events. Because a VMS can display a wide variety of characters in each character cell, dynamic messages can be created by manipulating the timing of the display of individual characters, or groups of characters. Simple effects that are quite effective for roadways include blinking text, moving arrows, and the cyclic display of a sequence of messages with delays between them. An example of the latter is displaying a repeating series of safety messages, such as "BUCKLE UP", "DRIVE 55 FOR SAFETY", and "USE YOUR SEATBELT". Message complexity, information acceptance rate, and driver attention span all must be considered when utilizing these features on high speed roads.

Two fundamental technologies, light reflectance and light emission, are used to form the individual dots that create the letters of the message.

Light Reflective Signs

Light reflecting VMSs consist typically of a matrix of mechanically changed dots. The individual dot can be a flat disk that is black on one side, and colored on the other, or a ball or cube that has color on one half, or a split flap that exposes a colored surface when opened. Other implementations consist of a multi-part flap that some vendors have utilized to implement a "white" character for daytime usage, and a "fluorescent color" character for improved visibility at night. This technique has been extended by one vendor to allow display of six different colors for each pixel. A variety of techniques have been used to improve the visibility of these signs, including internal illumination and retroreflective surfaces. Because the dots are mechanically moved, a finite amount of time is required to change the message displayed on the sign. Different vendor's implementations result in a range of timing characteristics. On the slow end of the spectrum, rates of 30 characters per second are typical. At this speed, a sign with three rows of twenty-two characters per row will require over two seconds to change its message. Faster character write rates are available, some capable changing the entire message in parallel, but tradeoffs of power consumption, dot inertia, overshoot, and flutter all enter into the dynamics of the implementation.

To provide stability during periods of power outage so that dots do not randomly change position and display "garbage" on the sign face, and to reduce power consumption, some method of latching the dots into a fixed position is normally used. A common technique is magnetic, where a small fixed magnet is attached to the shaft on which the dot rotates. The dot is changed from its "dark" state to its "bright" state with a pulse of an electromagnet, thereafter remaining stable with no power input required. This has the advantage that a message that was displayed prior to a power failure will remain on the sign face.

These signs have a proven field track record, with a generally high reliability rate. Individual dots are rated in the range of 100 million operations. However, it is not uncommon to find individual dots stuck' either "dark" or "bright", as a sign ages. The signs are fabricated for easy repair, with each character cell being quickly replaceable, and individual dots being repairable. The technology is easily scaleable, with character sizes ranging from 2" to 18" in height. A wide range of colors can be used on the "bright" side of the dot, with white or yellow being most common, but green, red, orange, gold, and others are becoming available. Because of the mechanical nature of this technology, a weatherproof enclosure is required. Cost of these signs is in the medium to expensive range, depending upon size, mounting, enclosures, and various options. For many agencies, these signs have been the "mainstay" of their VMS implementations.

By mechanically rotating the disk, ball, or flap with different colors on the surfaces, the dots on the surface of the sign form letters. The key advantage of this type of sign is the maturity of technology, and the long experience of their usage. Another advantage is the continued operation of the sign during a power outage, since the dots are bi-stable -- requiring power to change their state, but not to maintain them in a particular state. The disadvantages include limited visibility under some lighting conditions, fading of color contrast over time, and mechanical failures resulting in a "stuck dot".

Costs of these signs is a direct function of the number of characters on the sign face, and the attention to detail and quality by the manufacturer. Since this type of sign is electro-mechanical, operational experience and product refinement based on many years of development have an impact on long term reliability. Large signs (3 rows by 20 characters/row) range in cost from $50,000 to $90,000, including installation and commissioning. Small signs (3 rows by 8 characters/row) cost $25,000 to $50,000. Cost of the support structure (sign bridge, attachment to overpass, or roadside poles) is in addition to the basic sign cost.

A related type of sign is the changeable seven segment numerical display. This technology is useful for the display of variable speed limits. A sign may be fabricated in the form prescribed by MUTCD for a speed limit signal with the numerical digits formed by remotely controlled displays. This technique produces an easily recognized, variable speed limit that is less costly than a full VMS.

Another related sign is the rotating drum sign, where several faces of a rotating drum (or several drums) can be used to display one of several messages. These signs can be configured with the same size, shape, and letter fonts as traditional static signs. Further advantages are their lower cost when compared to a "dot matrix" sign, and mechanical simplicity resulting in higher reliability. Their prime disadvantage is the limited number of messages that can be displayed on a single sign. The drum sign has applications where a fixed message (such as LANE OPEN/LANE CLOSED) has to be displayed for portions of a day. Their use for incident response is limited.

Light Emitting Signs

The use of an active light source at each dot (or pixel) of a VMS produces a light emitting sign. The original light emitting sign is the incandescent bulb matrix. This type of sign provides good visibility, and is currently used in commercial applications. However, it has fallen into disfavor for roadway applications due to the low reliability and high maintenance costs from bulbs burning out. Another major problem is heat as a result of the high bulb wattage, and the resultant power consumption. Some agencies in warm climates have found that they have to limit the number of bulbs that are simultaneously ON due to heat rise in the sign enclosure. In general, these signs are not favored because of these limitations.

Current technology developments utilizing "solid state" lamps over the past several years have produced signs with high brightness, simple control, and long life. The light source in these signs is the light emitting diode (LED). Until recently, the brightness of the LED has not been fully adequate for bright daylight conditions. In particular, the "amber" LED, which is preferred for roadway usage, has been difficult to manufacture with the desired characteristics. Early LEDs suffered from variability in light output between "identical" LEDs, and aging effects which reduced brightness (often non-uniformly) over time. However, about three years ago these problems appear to have been solved, and the LED sign is finding acceptance in the field with many major manufacturers fabricating these signs.

A typical implementation utilizes a group of LEDs (on the order of 15) to form each individual pixel. This increases the brightness of each pixel, and averages any small differences between adjacent LEDs. These signs have a very fast turn on and turn off time, removing the problems noted above with the rotating disk type signs. Because of the physics of the semiconductor junction and wavelength of emitted photons, LEDs have a limited range of colors. Red is the most common color, but yellow is preferred for most roadway signage applications. Green is also commonly available. Combinations of different colored LEDs are being used to implement "colored" signs. The small size of the LED, coupled with computer type integrated circuits, can produce displays with large numbers of individually controllable dots for special effect applications. The long life of the LED, combined with the inherent simplicity of the design concept, should result in very good reliability. Actual field experience, as these signs are deployed in large numbers, will have to be gathered to verify this expectation. Cost of these signs is moderately expensive, but that should change as their usage increases.

Enhanced visibility is the key advantage of light emitting signs. The ability to mix various color light sources to produce differently colored messages is also useful. The biggest disadvantage of these signs is their requirement for continuous power, making them non-operable during power failures. If power failures are common, and the sign is critical to continued operations, some sort of back-up power is required.

LEDs have had some problems due to loss of light output intensity due to the aging of the light emitting active elements. Intensity reductions on the order of 50% have been observed after 30,000 hours of operation. A side-effect of this problem has been brightness differentials as a result of differing power-on times. This results in variations between different dots on the sign. Newer generations of LEDs appear to have solved these problems, with preliminary reports indicating either no intensity loss, or even a slight gain. This is based on initial testing, with long term field results not yet available. Another benefit of these newer LEDs is their increased intensity, allowing a sign to be fabricated with fewer LEDs per pixel (resulting in a lower fabrication cost), or a brighter sign with the same number of LEDs, or the ability to operate the LEDs at lower power (prolonging their life and reducing the aging effects).

Costs of LED signs is controlled by the size of the sign (number of characters on the sign face), the quality and reliability focus of the manufacturer, and the type of LED used. The newer, high-output amber LEDs are more expensive than older devices because of limited manufacturing yield and the need for the supplier to recover development costs. As with all semiconductor devices, component prices will decline fairly rapidly - especially as sales volumes increase. Large signs (3 rows by 20 characters/row) range in cost from $60,000 to $130,000, including installation and commissioning. Small signs (3 rows by 8 characters/row) cost $40,000 to $60,000. Cost of the support structure (sign bridge attachment to overpass, or roadside poles) is in addition to the basic sign cost.

Hybrid Technology Signs

The combination of a rotating disk or shutter in front of a light source produces a hybrid of mechanical motion and light emission. If the rotating disk is colored on one side, the light source "enhances" the message on the sign, providing additional visibility and "punch" for longer distance viewing. Some vendors consider this an enhancement of the basic rotating disk/shutter sign' while others explain their product as a totally different technology.

The LED is often used as the light source, with the LED being mounted behind the disk, and the disk serving as a shutter to permit the LED to be seen when the disk in the "bright" position, and masking the LED when the disk is in the "dark" position. One implementation mounts the LED off center, with a hole through the disk. When the "bright" side of the disk is visible, the hole is positioned over the LED. When the disk is rotated so that the "dark" side is exposed, the hole and the LED no longer coincide, and the LED is masked. Different vendors implement this same basic idea with a range of schemes, all effectively performing the same task.

A variation of this approach utilizes digital control technology that is connected to the circuit that controls the disk, and turns off the LED at each pixel when the "dark" side of the disk is exposed. This technique requires a location within each pixel that is constantly visible, and works well with circular dots where the LEDs can be located in a "comer" of the pixel. However, with split flap pixels that are square or rectangular in shape, the locations for mounting the LEDs are limited.

The approach of combining a light source with a light reflecting sign is an effective manner for increasing the visibility of the basic VMS, producing a good combination of daytime and nighttime usage. The prime reliability concerns are those of the basic sign. Cost is greater than that of the basic sign, and the performance enhancement must be considered within the constraints of the project.

A matrix of shuttered pixels, with each pixel containing a fiber optic bundle that is illuminated by a high intensity light source is another combination used by some vendors. The concept utilized with this design is that of a light source for several characters (on the order of three or more), and bundles of optical fibers to "pipe" the light to each individual dot on the sign face. One configuration utilizes a rotating disk as the shutter. In another configuration, the shutter is assembled with its rotational shaft perpendicular to the sign face. This shutter functions in a manner similar to that of a camera, alternately blocking or uncovering the light source. The mechanical orientation of the shutter, and its motion, seem to result in enhanced reliability.

The light source is a high intensity light bulb, similar to that used in a slide projector. The brightness of each individual dot is several times brighter than that obtainable with the hybrid LED sign. A useful design "trick" is to utilize two separate bulbs for each fiber bundle, with an automatic switch over circuit when a bulb fails. Monitoring the current flow of the small number of bulbs involved in this design is convenient, resulting in the ability to report a bulb failure to the central control station. The second bulb can also be used to produce an "overbright" condition for poor visibility conditions, such as fog. Another convenient feature utilizes a motor driven colored filter between the bulb and the fiber optic bundle to produce different colored characters on the sign face.

This type of sign has carried a higher price tag, making it the "cadillac" of VMS applications. The prime selling feature of these signs has been their brightness and resulting high visibility. Some vendors emphasize the reliability of their signs, which may be more a result of high quality manufacturing and engineering, than the fundamental technology. Competition, other market forces, manufacturing efficiencies, and related factors may eventually push the price down to being more competitive with other technologies. As more of these signs are installed and field experience gained, their relative merits will be more sharply focused.

The combination of devices (light source and mechanical shutters) used to create a hybrid sign increases the cost about 20% over either light reflective, or a light emitting sign. However, the increased visibility is a key benefit that is often required.

The cost of hybrid signs is also dependent upon the size of the sign (number of characters on the sign face), and the approach taken by the manufacturer. The "flip-disk" signs, to which LEDs or fiber-optic light sources are added as an enhancement cost 15% to 20% more than the basic sign. Thus, for a large sign (3 rows by 20 characters), the cost will be in the $60,000 to $105,000 range. And a small sign (3 rows by 8 characters) will cost $30,000 to $60,000. The fiber optic sign that utilizes shuttered pixels is primarily available in a 3 row by 18 character configuration, and costs about $135,000, including installation and commissioning. Cost of the support structure (sign bridge, attachment to overpass? or roadside poles) is in addition to the basic sign cost.

VMS Control Systems

As the number of individually controllable elements on the sign face increases, the complexity of the control requirements increases. For all but the simplest rotating drum signs with just a few messages, some sort of computer based control is required. The manufacturers have selected a variety of microcomputers to meet this need. A few manufacturers have selected the Model 170 intersection controller as the microcomputer, which has the advantage of utilizing a standard item of hardware that is familiar to highway agencies. In other cases, the vendor has developed a special purpose microcomputer for controlling the specific sign they manufacture. In all cases though, a unique software package has been developed for each implementation.

Similarly, the command set used for communication between these signs and a control location is unique to each vendor's system. This command set is called the "communications protocol".

For an agency getting started with VMSs, a fully packaged system from a single supplier is simpler because the vendor can be assigned total responsibility for the system. But the "proprietary" nature of each vendor's implementation (because standards have not yet been defined) creates difficulties when trying to integrate equipment from several vendors into an overall system. An agency can easily get "locked into" a single supplier, when there are superior or more cost-effective products available. Or the agency can suffer from poor support, or a product being "orphaned" when a newer model is introduced or a company is bought out.

In any application of VMSs where more than a "few" different messages are to be displayed, some form of central control and operator interface is required. The "central" control computer supplied by the vendor for remote access to and monitoring of the signs is usually a PC, but often with vendor specific hardware enhancements such as unique serial communications boards. The software that runs on the PC is unique to each manufacturer's implementation, and ranges from "convenient" tb "obtuse" in its user interface. Prices for the central system range from little more than the cost of the PC itself, to many times that, depending upon the features, the total system size, and the vendor's perception of the value of the central control system. The complexity of this software must not be underestimated. There are a great many features, interdependencies, database management issues, and operating subtleties to be handled, all of which contribute to the implementation difficulty and resultant cost.

The challenges associated with the control system can be addressed by carefully understanding the operational needs of the system, considering the growth requirements and future needs. In all cases, the vendor must be required to supply full documentation of all system components. The details of the communications protocol are especially important, so that existing signs can be integrated into a larger system when the agency's needs evolve. Another option that will be available in the near future is the specification of the National Transportation Controller/IVHS Protocol (NTCIP). This protocol is currently under development by NEMA/FHWA for NEMA/170 controllers, and will be extended to VMSs after the initial traffic controller work is completed. Selection of a VMS on the basis of ease of integration into a future larger system will usually be beneficial as the overall scope of this type of traffic information system increases.

VMS Communications

The connection between a VMS and the central processor can be provided by a standard serial data communications link. Data requirements for signs are usually small. VMS systems are often implemented with a precanned library of messages. An operator usually needs only to select a pre-composed message, resulting in a very small communication load. If a completely new message is typed in by an operator, the communication load is only slightly higher. A complex message with graphics will require a larger amount of data to communicate the display to the sign. The communication link to a VMS will not generally need to operate above 1200 bps. This data rate will allow roughly 120 characters per second to be transmitted.

When a secure "closed" communications system is required to prevent unauthorized access to VMS control capability, an owned or leased communication link is necessary. Although the public switched telephone network is an "open" system, security measures can be added. Security measures could include the use of encryption devices and/or call-back security. Encryption involves the transmission of messages in a code that cannot be easily reproduced with a personal computer. Call-back security involves the placement of a call to the VMS and entry of an identification code. The VMS then places a call back to the control point before allowing access to changes in the sign message.

2. Highway Advisory Radio

Highway advisory radio (HAR) is widely used to provide motorist information to travelers in a limited geographic area. Non-commercial information services include construction and taffic congestion information, possible alternate routes, traveler advisories, parking information at major destinations safety information, availability of lodging, rest stops and local points of interest. AM broadcast-band, low power level equipment has been used to provide this information on two frequencies, 530 KHz and 1610 KHz. Presently, the standard broadcast frequencies between 530 KHz and 1700 KHz are available, in 10 KHz increments, provided there is no interference with existing stations. The transmitters signal must also be low pass filtered in the audio range, to about 4 KHz resulting in a voice quality much like telephone transmission (between 3 KHz and 20 KHz the filter must attenuate at 60 log (f/3)dB where f is the audio frequency in KHz). The HAR transmitter consists of a device to record and playback messages, a radio transmitter, and an antenna. There are three different configurations used for HAR, vertical antenna, "leaky cable", and micro-transmitter. Regulations governing the operation of HAR systems are defined by the FCC rules in Part 90.242.

Vertical Antenna

Probably the most commonly used HAR system utilizes a vertical antenna. This type of HAR is termed a Traveler Information System (TIS) and must be appropriately licensed. A single vertical antenna produces an omnidirectional (circular) radiation pattern that diminishes uniformly as the square of the distance from the antenna, provided there are no geographical obstructions.

FCC regulations for vertical antenna HAR/TIS stations include the following requirements:

- A separation of at least 15 kilometers from the 0.5 milivolt/meter day-time contour of any AM broadcast station operating on the same frequency.
- A separation of at least 130 km from the 0.5 milivolt/meter daytime contour of any AM broadcast station operating on the same frequency.
- The height of the antenna must not exceed 15 meters above ground level.
- The RF output of the transmitter must not exceed 10 watts.
- minimum distance of 15 kilometers must be maintained from any other Vertical Antenna HAR/TIS station.
- minimum distance of 7.5 kilometers from a "leaky cable" antenna HAR/TIS at the same frequency.
- A frequency stability of +/- 20 Hz must be maintained.
- Signal field strength of antenna emission at the operating frequency must not exceed 2.0 milivolts/meter at a distance of 1.5 kilometers from the HAR antenna.

"Leaky Cable" Antenna

A specially designed lightly shielded coaxial cable is used to provide the antenna for this type of HAR/TIS transmitter. The signal transmitted from this arrangement is strong near the antenna, but dissipates rapidly when the distance from the antenna increases. Compared to a vertical antenna system, much more control of the emission field strength is available. There is less chance of interference with other radio services. Multiple HAR/TIS systems could be operated along a roadway with different messages for traffic in each direction.

FCC regulations for a "leaky cable" antenna HAR/TIS stations include the following requirements.

- A separation of at least 15 kilometers from the 0.5 milivoltimeter daytime contour of any AM broadcast station operat

- The maximum length of the cable antenna must not exceed 3 kilometers.
- The RF output of the transmitter must not exceed 50 watts.
- A minimum distance of 0.5 kilometers must be maintained from any other HAR/TIS "leaky cable" station.
- A minimum distance of 7.5 kilometers from a vertical antenna HAR/TIS at the same frequency .
- A frequency stability of 20 Hz must be maintained.
- Signal field strength of cable antenna emission at the operating frequency must not exceed 2.0 milivolts/meter at a distance of 60 meters from any part of the station.

Micro-Transmitter

Very low power HAR transmission is permitted by Part 15 of the FCC regulations without requirements for a license. The area covered by a micro-transmitter is usually defined by a radius of 0.15 to 0.4 kilometers although some manufacturers claim twice as much. Part 15 of the FCC code includes the following requirements:

- The lead length of antenna and ground may not exceed 10 feet.
- Any standard AM frequency between 530 KHz and 1705 KHz may be used.
- The RF output of the transmitter must not exceed 100 milliwatts,

Message Record/Playback

TIS messages to be broadcast on a HAR are usually recorded on an audio tape recorder and more recently in digital memory. Digital memory is preferred since it uses no moving parts, and thus does not require periodic cleaning or maintenance. Some devices offer features that include:

- message capacity of nearly half an hour.
- ability to retain messages during power failures.
- provide concatenation of various stored message sequences in any order to form the broadcast message.
- double buffer to allow playing one message while recording another.

Digital memory is available in several varieties:

- EEPROM (Electrically Erasable Programmable Read Only Memory)
- DRAM (Dynamic Random Access Memory) low cost but inefficient and sensitive to power fluctuations.
- SRAM (Static Random Access Memory) low power consumption can be battery backed-up with on-board lithium battery. Recent price drops make SRAM a good candidate for digital memory.

Transmitter

The function of the transmitter is to convert the audio signal from the message record/playback subsystem into a modulated AM radio signal to be transmitted from the antenna. Various classifications of transmitters are available. The power amplification stage of the transmitter is characterized by an alphabetic letter A through D to describe the linearity and efficiency of operation. Class A is the most linear and least efficient, while Class D is essentially switched on and off for various parts of the output signal and hence is the most efficient. Class D transmitters have a typical efficiency of 75%. Greater efficiency results in less heat losses and hence better operation. Efficient transmitters can be kept in sealed enclosures to protect them from dirt and moisture, thereby extending their useful life. Highly efficient transmitters will be more conservative in use of battery power during power outages.

Vertical Antenna Systems

It is desirable to place a HAR antenna in an area that has few obstructions to radio signals. Large buildings, geographic obstructions, trees, metal towers and overhead power lines should be avoided. An ideal site is a flat open field that is several hundred feet across. Good soil conductivity is another important factor. A radio ground plane can be improved with radials composed of heavy gauge copper wire buried about 12 inches below the surface and extending about one hundred feet in all directions from the base of the antenna. Ground rods are usually attached at the ends of the radials as well. Special chemical systems are available to provide a ground plane where available space may be as small as 10 feet in diameter.

The antenna must be tuned to the operating frequency. Both electrical and mechanical means are usually used to adjust the antenna and lead in cable to the transmitter output, to this provides maximum radiation from the antenna.

"Leaky Cable" Antenna Systems

Cable antenna systems are usually run in conduits and either suspended near the roadway or directly buried. A cable antenna is generally considered to be more expensive to install than a vertical antenna. If buried the antenna is easily damaged by roadway construction, roadside guiderail, sign and delineator installations as well as attack from rodents.

System Control

Most systems allow remote control that can be provided either from a touch-tone telephone or a personal computer. Telephone control is accomplished by interpretation of dual tone multi-frequency (DTMF) tones as commands, from a touch-tone phone. Some systems utilize voice prompts to instruct the operator to utilize the remote control features of the recorder and provide status messages. Under computer control, all functions and diagnostics can be controlled from a PC. The control software could incorporate a graphical user interface (GUI) to make system operation clear and intuitive.

Some systems allow the message to be composed and digitized at the PC before transmission to the HAR. The use of such a digital transmission reduces noise that might be introduced by this transmission. The resultant broadcast is clearer and more easily understood.

The communications link between the HAR site and the control point could be standard telephone, cellular telephone, owned cable, radio or fiber optics. Multiple HAR micro-transmitters could be utilized on the same frequency, transmitting the same message, provided that they are carefully synchronized. A fiber optic interconnect could be utilized to provide this means of synchronization.

Most HAR systems are able to operate in a mode that provides live message broadcast should the need arise.

Notification Signs

Signs advising drivers to tune their radios to the frequency required to receive a HAR broadcast should be placed near the edge of the reception area. Signs with flashing attention lights that are activated when an important message is being broadcast may prove useful to motorists.

3. Kiosks

Another medium for traveler information is the use of kiosks. Kiosks, in this instance, are-video screens that display maps and/or text information regarding traffic, incident and transit information. Placed strategically at shopping malls, schools or large places of business; kiosks can provide pre-trip information. This pre-trip information can be used by motorists to plan alternate routes around congested areas or around incidents. Transit users can plan alternate routes with information provided on the status of transit vehicles. Communications from the Traffic Operations Center and the Transit Dispatch Center to the kiosks is vital to the success of a kiosk system.

The Port Authority of NY&NJ is using kiosks in its airports as well as at the PATH (Port Authority Trans-Hudson) train platforms. The information provided includes time of arrival of next train as well as news and current event information. The NYSDOT provides kiosks at certain offices along the INFORM corridor in Long Island. A live map of traffic flows along the various roadways instrumented by the system.

4. Dial-In Systems

A useful pre-trip informational tool is the Dial-In System. A telephone number is established for the public to call for current traffic conditions for the ATMS. Usually, the messages are prerecorded with the time and date so the caller knows the age of the traffic information. This system could be set up as a toll free number or as a toll call. Once the call is placed, choices could be given to enter the highway route number or in the case of transit the bus line number. The recording would provide details as to traffic conditions at various interchange locations.

Information from the Traffic Operations Center and Transit Dispatch Center must be fed to the Dial-In System operator to update the recordings.

In New Jersey, both the Garden State Parkway and the New Jersey Turnpike have Dial-In Systems. The Turnpike provides a toll free number (I-800-33NJTPK) and provides choices for north, central and southern sections of the Turnpike. The Parkway's number is a toll call (l-908-PARKWAY), but also provides choices for north central and south sections of the Parkway. The New York State Thruway provides one toll free number (l-800-847-8929) for traffic conditions on the entire NYS Thruway between New York City and Albany, west to Pennsylvania, the New England Section, Cross Westchester Expressway, I-84, Berkshire Section and Niagara Section.

C. Traffic Signal System Architectural Alternatives

This section provides a brief but comprehensive comparison of the different traffic control architectures currently commonly available.

The types are:

- UTCS (Figure D-1)
- UTCS-Hybrid (Figures D-2 and D-3)
- Closed Loop (Figure D-4)
- Closed Loop Hybrid (Figure D-5)
- Distributed Systems (with and without on-street Masters) (Figures D-6 and D-7)

UTCS, is now generally considered to be an outmoded concept, The UTCS-Hybrid concept refers to systems in Raleigh, N.C., Pinellas County, FL and Washington, DC, in which the UTCS software concept is used but the concept has been expanded to permit downloading and uploading of the controller database. The last two names may not be commonly recognized. That is, some users or vendors may call these architectures by some other name.

A closed loop hybrid system, by definition, is a closed loop system with external modems. Theoretically, at least in New York, for a system to be called a closed loop system, the on-street master, the coordination units and the controller must all be manufactured by the same company. Since, in the closed-loop hybrid systems, the controller can be of separate manufacture, it is not within New York's definition of closed loop.

Distributed systems also may have other names. In this definition, a distributed system is one which uses most of the closed loop philosophy, but has modified that approach somewhat so that controllers in a traffic control section and on a communication channel need not be synonymous as in a closed loop system. That is, timing plans, timing plan schedules, reports, and global commands can be issued without regard to which channel a controller is assigned. In this respect, the distributed system is closer to a UTCS system than it is to a closed loop system. On the other hand, the utilization of smart coordination units in the field which store the timing plans, schedules and do the once per second timing is identical to the closed loop philosophy .

1. UTCS

Intent of UTCS

UTCS (Urban Traffic Control System) was developed (starting in 1970) by the Federal Highway Administration in an attempt to standardize on a traffic control system concept. By doing so, it was hoped to reduce costs by producing a single software package applicable to all systems and which was in the public domain. By writing it in a universal language, it was then believed that hardware costs could be minimized since the software could work on many different computers. In addition, it was believed (rightly) that state-of-the-art hardware improvements could readily be incorporated into system designs.

The original system was developed in assembly language for an obsolete Xerox computer. Later versions were written so that they could be made portable; i.e.. could be run on more than one vendor's computer. The early UTCS system software was limited in capability since:

- It was not designed to operate with Time Division Multiplexing (TDM) communications

- It was not designed to operate with actuated controllers

- It was limited in size to some 256 controllers.

Limitations of communications capacity (brought about by cost considerations) precluded bringing vehicle and pedestrian calls to the central computer. Therefore, there was no need to provide the logic necessary for actuated operation in the central computer. All of this logic remained at the street level.

The high cost (at the time) of solid state electronics thought to be capable of surviving in the environment of a controller cabinet dictated that an electro-mechanical time clock be used to provide backup coordinated timing.

The basic philosophy which dictated the design was the belief, at the time, that:

- Solid state electronics would never develop to the point where it would stand up to the hostile environment found in an un-air conditioned controller cabinet located at a typical urban intersection.

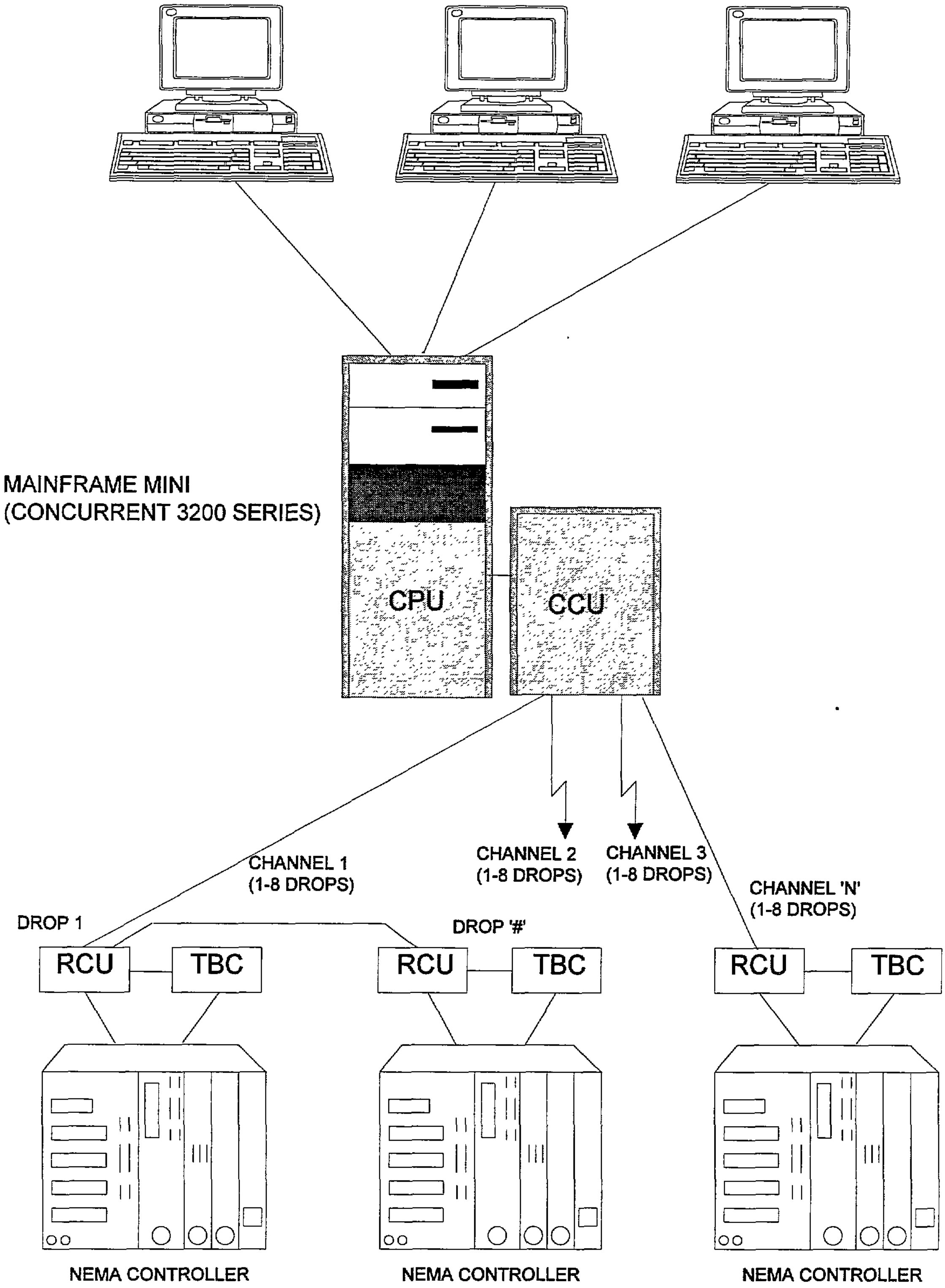
- TDM communication approaches, then very expensive and used mostly by the military, could never be made low enough in cost to be practical in a commercial application.

Thus, the design was based on keeping all complex digital computations at a central location where it could be in an air-conditioned and protected environment. It was believed at the time that the street control equipment would be an electro-mechanical device, much like a 3-dial, 3-offset controller. This device would provide the basic switching unit that would be commanded from central. In addition, it would serve as a very good back up system in case of central or channel failure.

UTCS-Type Architecture

Figure D-l illustrates the architecture of the UTCS system. The central computer (CPU) communicates to the street via the Central Communications Unit (CCU) then to the Remote communications Unit (RCU). The Remote Communications Unit decodes the commands from the central site and encodes the monitored messages from the controller assembly. The RCU also pre-processes detector data before it is uploaded to the central computer. The RCU also communicates with a Time Based Coordinator (TBC) which provides backup capability in the absence of data from the central site. When the RCU's watchdog indicates that there is no data forthcoming from central, it switches control to the TBC. It also contains a mechanism for delaying the transfer back to central control until it ascertains that the communication channel is solid; i.e., this precludes random switching back and forth between the two timing sources in the event of erratic communications. The controller shown here is a typical NEMA controller. However, a Type 170 controller can readily be included. In addition, pre-timed controllers can also be included.

Figure D-1



UTCS SYSTEM ARCHITECTURE

Traditionally, the number of drops that can be included in a channel is limited to 8. The number of channels that can be included in any system is limited by the size and speed of the computer. The drop limitation of eight (8) was established many years ago when the data rate of lines was limited to 1200 bits/second. With data rate availability to 2400 bits/second (on copper) and virtually unlimited (on fiber), the number of drops can be expanded with little difficulty.

Figure D-l shows graphics style multi-user interface terminals. Such usage is limited in UTCS systems at present. However, there is no technical reason that this concept can not be implemented in existing UTCS systems.

Design Limitations Caused by State-of-the Art Limitations

Absence of Actuated Logic

At the time of the UTCS system development, communication costs were projected to be much higher than they are today (because the techniques of the day required many more channels) and because of limitations of the computers of the era, the UTCS design paid scant attention to the implementation of actuated operation. Pedestrian and vehicle detector actuations were not brought back to central due to prohibitive communication cost) and logic was not added in the central software for implementing actuated operation (computers did not have sufficient power for the added logic). Therefore, by default, actuated operation was implemented in complex devices on the street. Thus, the seed was sown for moving more computational complexity to the street. This trend is still continuing.

The essential difference between a UTCS system and the new systems is that the "once-per-second" processing is done in the field hardware, including actuated operation, not in the central hardware. Typically, in a UTCS system, about 50% of the central computer's work is in doing once-per-second processing. All of that, in the more modern systems, has been moved to the field units.

User-Interface and Database Management

At the time of the UTCS development, user-interface to the system was accomplished through monochrome terminals and 80-column "IBM cards". Little was available at the time in advanced database management techniques. Computer memory was very limited; extensive use of overlay techniques was required. As a result, database loading was usually done with the traffic control system off-line; i.e., the system had to rely on backup system timing whenever databases had to be changed.

Monochrome terminals, used at the time as the primary interface between the computer and the computer system engineer, were modified to make them more of a user-interface device. The monochrome terminals were "dumb". They had virtually no computing capability or database storage capability. Computer graphics were unknown. Intersection and section-wide displays were only through electro-mechanical “maps” which used incandescent lamps to indicate intersection and detector operation.

The more modern systems, of course, take full advantage of today’s computer graphics, database management techniques, inexpensive memory, smart PC-based terminals, etc.

Because there is no competitive edge to be gained from investing in it, the private sector has made no attempt to improve the database handling techniques or to add color or color graphics capability to the basic UTCS system. As a result, the UTCS approach is often referred to as a dinosaur. However, those who make this statement tend to indict the whole system because of its lack of user-friendliness. It would not take much effort to bring the UTCS system’s user interface techniques up to the same standards as found in the new traffic control systems.

UTCS Algorithms

UTCS work was responsible for the development of the control algorithms which, for the most part, are used by all of the system architectures available today. The algorithms used to smooth/average data are about the same, the methods for traffic responsive operation have changed very little, Critical intersection Control, if used, still relies on the UTCS-developed algorithms.

The Federal Government made several attempts at improving traffic control algorithms. Although the attempts were well-meaning, the limited computer capacity of the day doomed these attempts from the beginning. Unfortunately, there has been little effort in the U.S. to improve upon this situation, even though current computers can now provide the computing capacity required.

Fortunately, however, this picture appears to be improving. The SCOOT/SCAT systems of the British and Australians are getting increasing attention. The Federal Highway Administration is undertaking programs implementing these approaches. In addition, a new program for the development of a new, US based approach, is being undertaken by the FHWA. If successful, this system might be available in the late ’90s. A whole array of new algorithms is discussed in a subsequent section.

Advantages and Disadvantages of the UTCS Approach

Disadvantages

The usual criticisms of the UTCS approach fall in two areas: High communications costs and difficult user-interface.

- The high communications cost critique is in comparison with the closed loop/distributed type of system. Because the newer systems have moved the once-per-second communications to the street, it is no longer necessary to transmit data to the street every second. As a result, the data bandwidth required per intersection is lower so that more drops can be added per channel. The comparison between UTCS and the distributed systems is at a ratio of 3 or 4 to 1. Although the cost reduction due to the increase in the number of drops per line isn't as dramatic as this ratio would indicate. there is some reduction in cost when either leased telephone lines or user-owned copper lines are used. More significant cost reductions may occur when considering radio or alternate communication approaches.

The difficult user-interface critique is justified. However, a concentrated, central effort at developing a uniform database management system and user-interface system, applicable to all current UTCS systems would readily bring these UTCS systems out of the dinosaur era into the modern era. Since the newer systems offer virtually no other advantage over the UTCS approach, this would provide many user's with improved operation, at low cost, to tide them over until systems with advanced algorithms can be developed and implemented. This will probably be at the start of the century.

- Inability to download or upload the controller database (yellows, all reds, minimum greens) or backup system timing database (cycle, phase splits, offset). Some UTCS systems can download the backup system timing database. At least one UTCS system (Washington, DC) can upload/download the controller database - it uses Type 170 controllers, throughout. Techniques are being considered for doing this with new NEMA controllers, but it is quite difficult when the individual manufacturer's protocol is unknown. A protocol standard has been proposed by Computran. This standard; Protocol 90" is an attempt to unify various manufacturer's communications to one design. It has been adopted by several manufacturers.

Advantages

The UTCS approach has several advantages over the other systems:

- With respect to the Closed Loop Systems, all database and operation activities are with respect to "traffic control sections", not with respect to "channels". Much bigger systems can be implemented without the major problems caused by trying to coordinate multiple closed loop systems. This advantage does not hold in comparison to the distributed system however.

- Much lower complexity of street equipment required permitting lower cost, higher reliability, smaller controller cabinets, etc. Unfortunately, there is perhaps only one known system which takes advantage of this approach; that one is currently being installed in the Borough of Manhattan, New York City.

- Nothing in UTCS is proprietary. It can mingle NEMA controllers, pre-timed controllers, Type 170 controllers, etc. in a single system.

- Since the algorithms are stored and used in one central site, modified algorithms can be inserted, tested and utilized as they become available.

- All data from the street is instantaneously uploaded, most notably log data, including alarms.

- High resolution of the volume and occupancy data; resolutions at the one-second level as opposed to the other system's once-per-cycle, once-per-minute or once-per-5 minute levels.

Other factors Leading Away from the UTCS Concept

Backup System Operation

In most current UTCS systems, a backup system is involved. A backup system requires that, in the absence of any information from central, the controller still must provide some coordinated operation level-of-service, even if it is relatively poor in comparison with its performance under central control.

This backup system is provided by some form of TBC (Time Based Coordinator). This device provides some sort of time clock control of NEMA-type commands which cause the controller to advance through its intervals and phases, even when the central computer is not being heard from.

It soon became obvious to some of the system vendors that there was a certain redundancy between the backup system and the foreground system. That is, if the system has the capability of storing and implementing backup timing plans at the street level, then why continue to compute and implement the timing plans from central. This thinking led to the closed loop and distributed system concepts.

Non-Proprietary Controllers and Communications

In addition, in most UTCS systems, the controllers are standard NEMA types and any manufacturer's version will work with them. The only device which is universal is the communications interface device. However, since communication protocols in UTCS systems are not proprietary, any vendor can manufacture the communications device.

This property of UTCS systems led various old guard vendors to invest heavily in new approaches. The new approaches emphasized proprietary techniques which would guarantee that the vendor, once "having his foot in the door", could continue to expand his system without competition.

2. UTCS -Hybrid Systems

Intent of UTCS-Hybrid Systems

The UTCS-hybrid system is an attempt to combine the advantages of the UTCS system with the ability to upload and download the controller database. Some of these have been implemented in Raleigh, N.C., Washington, DC and Pinellas County, FL.

UTCS-Hybrid Type Architecture

Figures D-2 and D-3 illustrates the architecture of two different UTCS-Hybrid system configurations. In either case, the central computer (CPU) communicates to the street via the Central Communications Unit (CCU) then to the on-street unit.

The two system architectures from that point on are quite different. The Washington system utilizes Type 170 controllers throughout. By standardizing on this unit, in which the user can define the communication protocol as well as all of the rest of the software requirements, the communication system transfers data directly from central into the memory of the 170 controller. Normally, control is through the usual NEMA functions, as in a standard UTCS system, certain messages, however, can be directed to the 170 controller's internal memory.

In the Pinellas system, the jurisdiction's intent was to provide the protocol to the all potential controller vendors. The vendors would manufacture their units so that they could utilize the protocol as established. The jurisdiction hoped that they would be able to get at least three vendors to join in so that their could be a reasonable hope for competition as the system was expanded. Unfortunately, to this date, only one vendor has responded.

All of the rest of the system operation is comparable to a standard UTCS system.

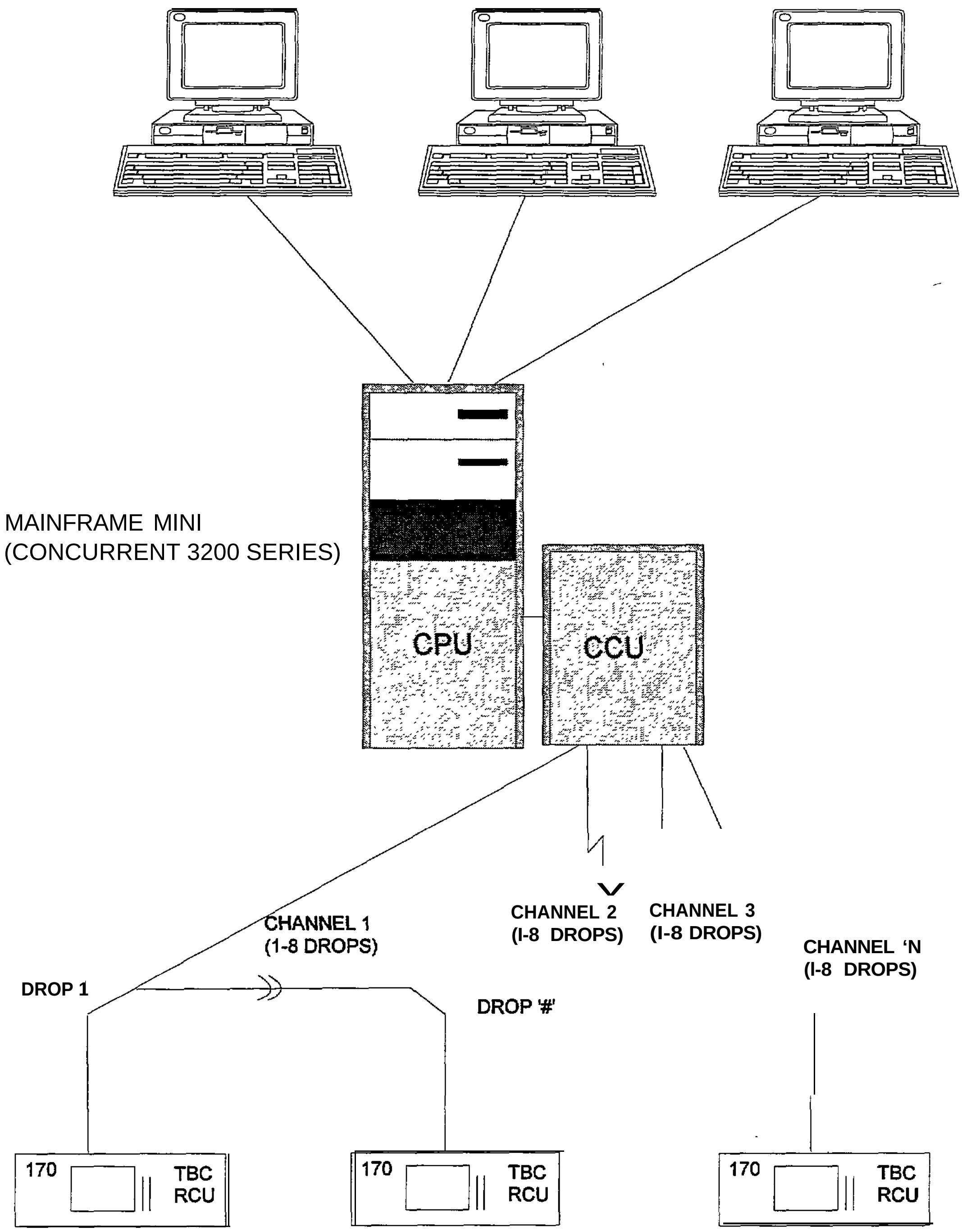
Advantages and Disadvantages of the UTCS-Hybrid Approach

Disadvantages

- All of the usual disadvantages of the UTCS system are maintained in the UTCS-Hybrid approach with the exception that the controller database is uploadable and downloadable.

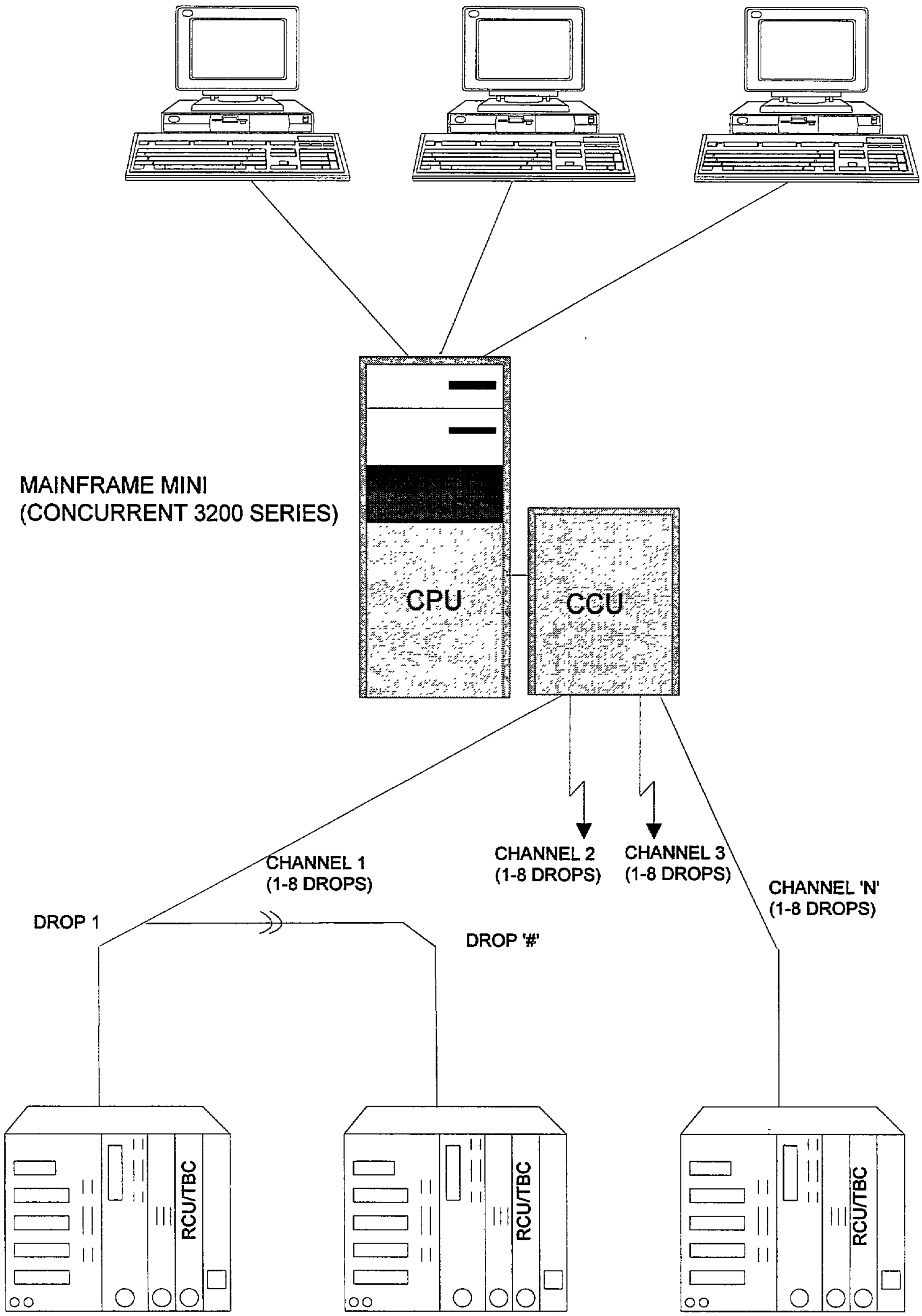
- Because of the system's limited bandwidth, only a few bytes of message data can be downloaded/uploaded each second. Thus, some messages can take considerable time to upload/download .

Figure D-2



UTCS-HYBRID SYSTEM ARCHITECTURE (WASHINGTON, D.C.)

Figure D-3



UTCS-HYBRID SYSTEM ARCHITECTURE (PINELLAS COUNTY)

Advantages

All of the advantages of the standard UTCS system applies in this configuration. In addition, the controller database is uploadable and downloadable.

Overview of Modern Candidate Architectures

State-of-the-Art Changes Since the Early UTCS Days

When the LJTCS system was originally developed, actuated controllers were in the minority. Almost no consideration was given, at that time, to bringing back local "calls" (pedestrian or vehicle). Actuated operation could not be commanded from the central computer. Because the UTCS concept (except for actuated operation) minimized the cost and complexity of the local controller, the controller manufacturers fought back. They developed and sold actuated operation (which UTCS could not do from central) to the traffic community. Although at first, actuated controllers were very expensive, the rapid expansion of solid state electronics with its increasingly reduced cost and increased reliability, soon made actuation at every intersection very close to a mandatory requirement.

Since the UTCS system was a Federally developed system which was turned over to the States for implementation, then abandoned by the FHWA, at no time was any consideration given to adding actuation capability to the central software. In due time, the concept of the Time-Based Coordinator (TBC) developed. The TBC was originally designed to provide coordinated signal timing at locations where communications from a central point was too expensive or (at least) not cost effective. A TBC provided a mechanism for storing several timing plans and a timing plan scheduler in a separate box which interfaced with a NEMA controller through the HOLD and FORCE OFF mechanism standardized by NEMA.

The traffic engineer now had a tool for controlling signals without a central system. However, the TBC, without a central system, did not allow the traffic engineer to monitor the signal nor to alter a timing plan or a timing plan scheduler without going to the field to do it. In addition, regardless of how well designed, TBCs had tendency to drift in time with respect to an adjacent TBC. This, of course, hurt the concept of coordinated operation using TBCs, alone. In spite of this concern, there are many coordinated systems installed which rely solely on TBC operation. The TBC did provide capability for backup not previously available. TBC manufacturers standardized (to some extent) on being able to store 32 timing plans in the TBC in addition to the necessary daily schedule, weekly schedule, holiday schedule and a special events schedule.

Thus, a mechanism was now available to store and implement a wide variety of primary and back up timing plans at the street level, relatively independent from a central source, leading to the closed loop concept.

3. Closed Loop System Concept

Intent of the Closed Loop concept

Closed loop systems developed in response to a general demand for arterial systems which would provide coordination on a fairly isolated arterial. The UTCS system concept, because of the high cost of mainframe machines, was over-priced for this application.

The Closed Loop System's development corresponded to an upgrade of the use of time based coordinators. These devices, as stand-alone units, stored many timing plans and timing plan schedules. Their operation included computation, with one-second resolution, the time force off and yield information that was to be sent to the controller. In addition, their time keeping mechanisms were adequate to maintain relative timing among a group of such devices to better than one second. Thus, coordination between adjacent intersections could be maintained, even after a timing plan was changed; as from AM peak to MD peak.

The next natural progression of this device was to provide a means by which data could be uploaded so that the system could be monitored at some central site. Since the design concept concentrated on independent and remote arterials with a limitation of between **4** and 30 contiguous signals, it was found that full time monitoring was not required. Therefore, the concept of monitoring, from central, the operation of each closed loop system only when needed seemed a natural one. Therefore, a master device, located in the field, was developed as a central communications monitoring point for each system. A single central computer, along with auto-dial modem techniques, was developed so that the field master could call central (or be called by central) only on an as-needed basis. It soon became obvious that sending commands to the system from central, via the field masters, was also an essential element of the system.

At this point, several manufacturers, almost in unison, came up with a system concept that:

- Utilized the Time Based Coordinator as the primary storage and computing element of the system
- Combined foreground and backup timing in the one unit; i.e., abandoned backup timing altogether.
- Negated the need to transmit YIELD/FORCE OFF data to the central with a once-per-second resolution. This reduced communication bandwidth requirements, thus communication costs.
- Filled a need for a small arterial-oriented traffic control system; a need that UTCS was considered to be too large and to expensive to fill.

- Controller timing functions such as yellows, all reds, minimum greens, etc., can be changed without physically visiting the field site.
- Because once-per-second data need not be downloaded, the data bandwidth for each controller is much narrower than in a UTCS system. Generally, vendors have chosen to permit up to 24 or 32 controllers (drops) to be placed on any channel.
- Since data need not be downloaded once-per-second, alternate means were found to avoid uploading data once per second.

Basic Operation of the Closed Loop System

In general, the Closed Loop System concept offers little advantage over a UTCS concept in terms of the method that timing plans are generated or are transmitted to the street.

The timing plans and the scheduler are stored at the local controller. The TBC acts as the master clock and the master scheduler. As long as the clock in each TBC (in a coordinated section) remains accurate (to within one second) and as long as the timing plan and the timing plan schedulers all remain consistent, coordinated operation is maintained. Controller transition timing is computed in the TBC.

As stated previously, most controller manufacturers permit the storage of 32 timing plans in their TBCs. At least one stores timing plans as a cycle length, split sets and an offset. The split sets and the offset are stored as percent of the cycle length. Therefore, there is an almost unlimited number of "timing plans" available.

Since the UTCS system had a traffic responsive mode, closed-loop system vendors included it in their systems. Because a "field master" was already available for a number of purposes, the means to compute traffic responsive timing plans is included at that point. Traffic responsive operation requires the analysis of data from several system sensors located at several points within the section. All of the data from these sensors must be available at one site central to the section. Computations are made on this data for the purpose of selecting which of the many timing plans available would be best to satisfy the requirements of the traffic demand determined to be available by the system sensors.

Different manufacturers mechanize the traffic responsive calculations in slightly different ways. However, the technique is essentially the same. That is, the data available from the system sensors are smoothed or averaged and the resulting processed data is compared against threshold values. These threshold values, in turn are related to specific timing plans. The timing plan selected for implementation is found from these threshold comparisons. The timing plan "indicator(s)" are downloaded to each coordinating unit (TBC). The coordinating unit then implements the selected timing plan.

Event Logging

The closed loop systems maintain a log of events that occur. The log is maintained in each TBC. At some point, the log is transferred to the field master where it is stored for some period of time (or until a given number of events are stored). If not uploaded to the central processor, the event log might be lost, or at least a portion of it might be overwritten by events occurring later. The events logged include failures, timing plan changes, special function actuations, power outages, etc.

Data Collection

Closed loop system vendors emulate the UTCS system's ability to collect traffic flow data, to display it and to archive it for further analysis. In addition, the collected data is used for traffic responsive operation.

Because there is no need for second-by-second data transmission to the street, the closed loop manufacturers do not provide second-by-second monitor data. Therefore, they do not provide volume and occupancy data with the same degree of resolution available in a second-by-second system. In the current traffic control concepts, all of the volume and occupancy data are averaged, smoothed and otherwise filtered before use. There is, in reality, no need for data resolution to the one second level in the current system. Future traffic control algorithms, however, may require once-per-second data resolution.

Note that all of the features of closed loop systems described in this section also apply to the distributed system concepts discussed in later sections.

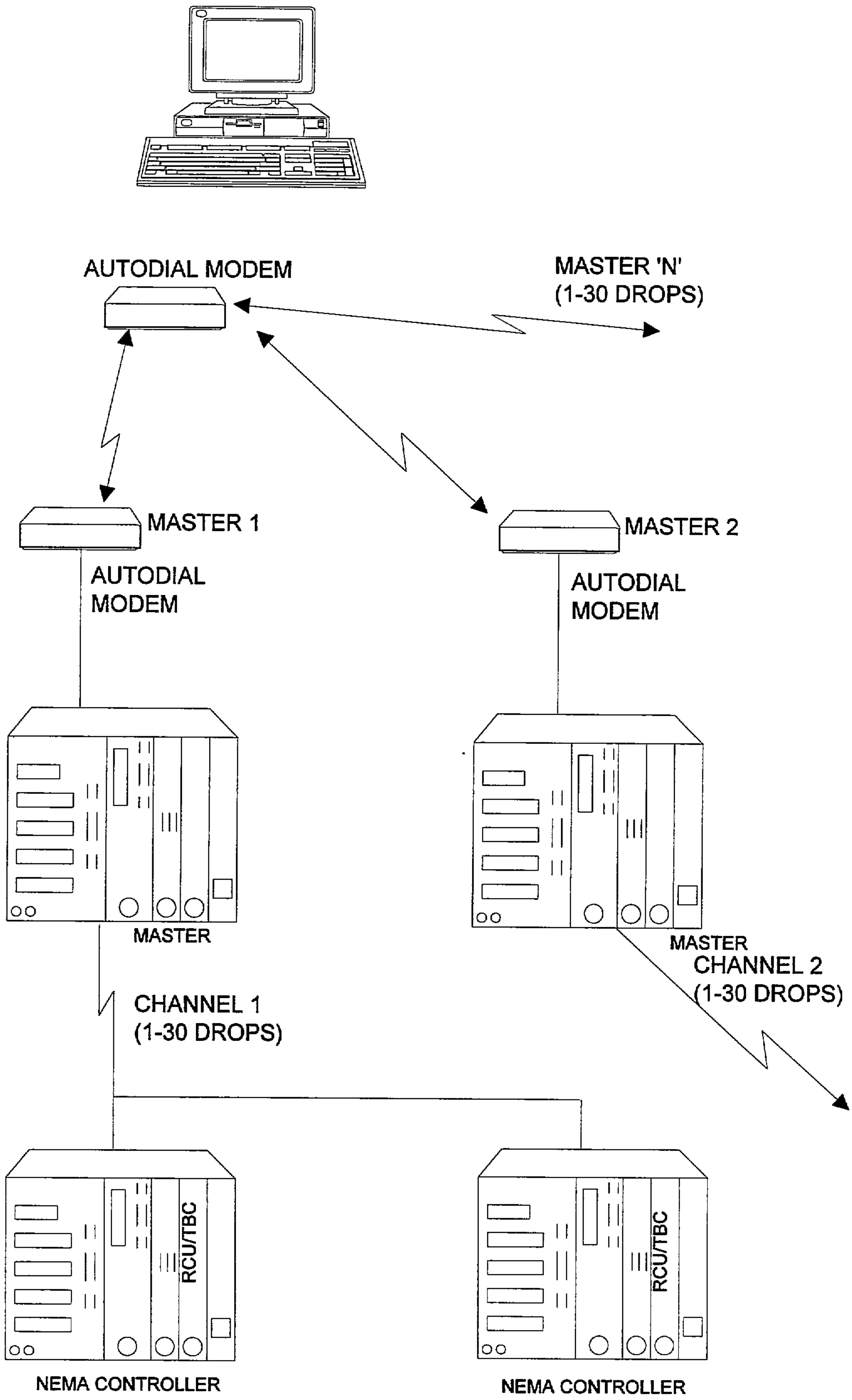
Closed-Loop Architecture

Figure D-4 illustrates the closed loop system concept.

Note that the central computer is merely a PC; it serves only as the user interface and data storage device. It provides no real-time computing capability for the system. A single Hayes-type modem is used to dial to the field or to accept calls from the field. A comparable companion modem is used at each field master. The field master does nothing much more than pass data from/to each field unit to the central computer. However, it also functions as a data storage device (until the central computer is available to upload the stored data) and as a computer to process traffic responsive algorithms.

The field master contains a built-in modem (usually a proprietary modem, not a Hayes-compatible modem) which communicates with a comparable modem in each controller cabinet. The modem (or RCU) is a plug-in module in the manufacturer s controller. In addition, the TBC (or in this variation, it is usually called a "coordination unit") is also a plug in module.

Figure D-4



CLOSED LOOP SYSTEM ARCHITECTURE

Because these units plug into the data bus of the controller, the modem/RCU has access to the controller's database as well as to the coordinating unit's database. As a result, anything stored in the controller can be uploaded (monitored) by the central site and can be modified (downloaded) from the central site.

Design Features

Restriction to Single Traffic Engineering Section

The basic closed-loop system design was predicated on the fact that each closed loop system would correspond to one arterial, This implied that a timing plan would be applied to the whole section; that there would not be other sections assigned to that arterial. It also implied that only one channel would be utilized between the field master and each local coordination unit.

Therefore, the design makes it very difficult for a single closed loop system master to provide coordinated timing for two different traffic control sections. For instance, suppose that after the installation of a closed loop system on an arterial, traffic conditions changed on the arterial so that it would be desirable to have different timing conditions at two different locations on the arterial. In a UTCS system, this would be done by making it two sections. In a closed-loop system, two sections can not be emplaced in a single field master. However, it is possible for some signals to have one cycle length; other signals to have a different cycle length. It is possible to provide the two section concept in a given master. Even more difficult is the ability to provide traffic responsive computations for each of the two sections. Since the closed-loop master only has one section defined in its software, it effectively can not select different timing plans for each of two sections.

Closed Loop System Algorithms

The closed-loop system vendors did little to improve the algorithms used in the UTCS system. The majority of the vendors use the standard UTCS algorithms for imposing traffic responsive timing plans.

Advantages and Disadvantages of the Closed Loop Approach

Disadvantages

- Difficulty in handling more than one traffic engineering section is the major difficulty in a closed loop system.
- Since the time base must be maintained equal by all of the controllers in a section, this becomes a critical factor in the maintenance of coordination.

- Log data and sensor data can be lost if for any reason, the central computer does not ask for it to be uploaded before it is overwritten by later data. The amount of data lost depends upon the size of the data buffer and how often it is uploaded.

- Alarm data can not be readily uploaded when the central computer is monitoring a field master other than the one in which the alarm occurs. Generally, it requires a one-to-one correspondence between field masters and central units to guarantee that alarm data is uploaded immediately.

- Volume and occupancy data are collected and stored with a resolution measured and averaged in terms of 1/2 minutes, minutes, or cycle lengths, rather than seconds. This may not be a problem at the present time; it may be a problem later as new algorithms are developed.

- The communication system and all elements of the system are proprietary. A master cannot talk to any other vendor's central, nor can it talk to any other vendor ' s coordination unit.

- Since all of the algorithms are stored and computed in the field unit, any algorithm

- change requires changing out the PROM module in one or more units at all of the

- locations in the system.

- Complexity of on-street equipment is high. This does not appear to constitute a problem since modern solid state electronics operates very well in the hostile street environment.

Advantages

- Communication costs can be reduced since communication bandwidths are lower. The amount of saving depends on the type of lines, the local tariff arrangement, etc.

- User interface capabilities are much enhanced, especially with respect to the older UTCS system approach

- Controller databases can be uploaded and downloaded. This can result in a reduction of the need for field site visits to modify yellows, all reds, minimum greens, etc.

- Central computer costs are much lower - generally a PC type of machine is all that is required.

- No backup system is required. The basic control concept and the backup control concept are identical.

Factors Leading Away from the Closed Loop Concept

Non-Proprietary Controllers and Communications

Unfortunately for the equipment users, each closed loop system vendor considers his communication protocols as proprietary. At one point in controller development, the NEMA organization agreed to share a common protocol (through the use of standardized A, B and C connectors). The UTCS system was able to communicate with all NEMA controller manufacturer's equipment because of this standard early "protocol". However, more recently, the NEMA group has not been able to agree on a standard protocol or on the desirability of making their protocols accessible to all. It is, therefore, impossible for one manufacturer's master or central computer to talk to another manufacturer's master or controller. Once a City has invested in one manufacturer's closed loop system, that city is virtually destined to sole source procurements of that manufacturer's equipment in all expansions.

Under State of New York procurement procedures, this is a difficult limitation. The most likely consequence is that a user may find that he has two or more manufacturer's systems in his central control room simultaneously. The major problem with this is that the user's system operators will be required to learn how to operate two or more systems; they must learn to prepare databases for two or more systems with different data requirements; they must learn to analyze data displayed or printed in different formats and in different units and at different time periods.

Another downside to the Closed-Loop system approach is its low data rate approach. Although this approach keeps communication costs down it may not be able to support the data requirements of future generation traffic control algorithms.

The inability to access two traffic control sections at one time led, in part, to the development of the Closed Loop Hybrid System. This system permits a single central system to communicate with all NEMA and Type 170 controllers. The downside, however, is the inability to upload and download the controller database (minimum green time, all red time, yellow time, etc.)

Limitation of Control Strategy Flexibility

The major disadvantage, to date, of the closed loop system is its basic inability to impose timing strategies across the boundaries of two or more masters. Closed loop systems on two arterials which are spaced 1/2 mile or more apart work well; the addition of one or more controlled intersections on links connecting the arterials creates a network; and a closed loop system can not handle network control very well.

4. Closed Loop Hybrid Systems

Intent of the Closed Loop Hybrid System

This type of system evolved from those vendors who did not wish to pursue the development of a controller. The system has a major advantage to the user (over a closed loop system) in that, except for the TBC and the RCU, he is not locked in to one manufacturer's hardware system.

It is not clear whether the advantage of a somewhat more open purchasing arrangement will override some of the cost disadvantages. The NEMA controller manufacturers, by using slots in their controller, can generally maintain lower costs than can a vendor who must supply a stand-alone shelf-mounted unit. In addition, the inability of the system to interface with the controller's data bus provides it a certain, albeit limited, technical disadvantage.

Closed Loop Hybrid Architecture

Figure D-5 illustrates the closed loop hybrid system concept.

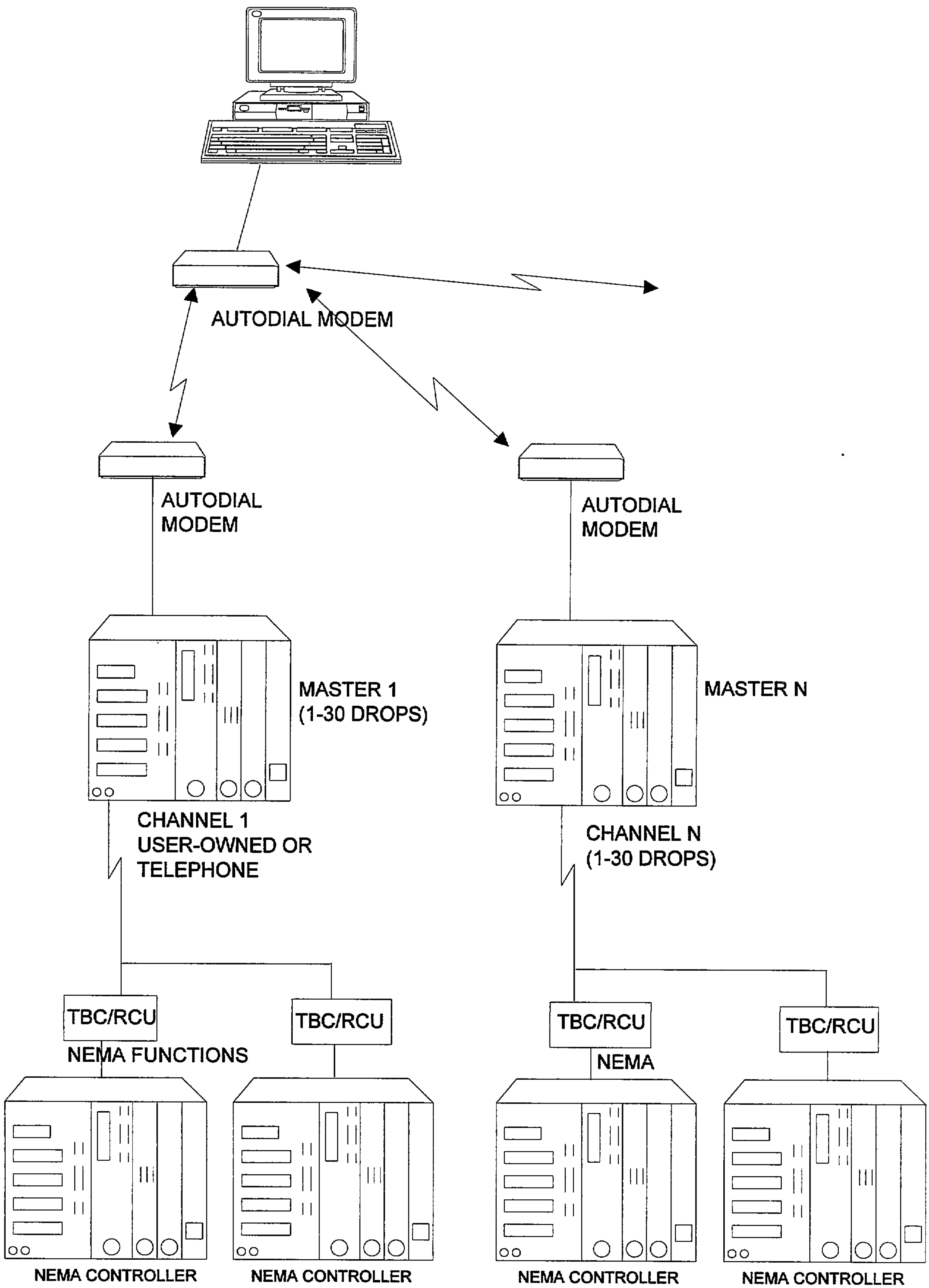
Note that the figure is identical to that of Figure D-4 with the exception that the TBC/RCU is shown as external to the controller. In this architecture, which is preferred by system vendors who do not manufacture a controller, the TBC/RCU drives the controller via standard NEMA hold, force off and yield commands. Because these units do not plug into the data bus of the controller, the modem/RCU has no access to the controller's database (although it does access the coordinating unit's database). As a result, nothing stored in the controller can be uploaded (monitored) by the central site nor can it be modified (downloaded) from the central site.

Design Features

Similarity with Closed Loop Systems

The closed loop hybrid system is almost identical to the closed loop system. The one diierence is that it can not down load or upload the controller database. Each controller cabinet contains the manufacturer's Remote Communications Unit. This unit is quite similar to the RCU used in a UTCS system. The input received from its master is command information (such as the choice of timing plan, timing plan parameters downloaded to it from central via the master, daily, weekly, yearly and special event schedules downloaded from central via the master, etc. Its output data uploaded to the master, then transmitted to the central, includes a log of all important events that

Figure D-5



CLOSED LOOP HYBRID SYSTEM ARCHITECTURE

transpired at the intersection, an accumulation of volume and occupancy data from each of the system detectors assigned to the local RCU, and alarm data which is sent to central on an expedited basis.

Advantages and Disadvantages of the Closed Loop Approach

Disadvantages

The major disadvantage of the Closed-Loop hybrid approach is

- The controller database can not be uploaded and downloaded:

Advantages

The advantages of the Closed Loop Hybrid System are, for the most part, the same as for the Closed Loop System. An additional advantage is:

- Any NEMA controller can be installed by the user. The only proprietary item is the RCU. It is proprietary because of the communication protocol - the identical situation that exists with the closed loop system. However, the RCU is considerably less expensive than the controller, cabinet and auxiliary equipment required when a new controller is purchased. Thus, the lack of complete openness is subject to less cost escalation than a complete controller assembly.

5. Distributed System Concept

Intent of the Distributed System Concept

This type of system was evolved by those vendors who saw a market position requiring a system with capabilities above that of a closed loop system, but who saw that the closed loop architecture had certain advantages over the UTCS architecture.

The major advantage of the approach, over the closed loop system, is that the traffic engineer does not have to be concerned with which controllers are assigned to a channel. All of that is transparent to him, just as it is in UTCS. In addition, any large jurisdiction which is considering a traffic system, but which prefers the closed loop system’ s user-interface advantages, may find this approach to be an optimum combination of the two approaches. A typical method for doing this is to provide a dedicated central computer which, in conjunction with the masters, can provide control across one or more masters. In effect, this type of system overrides the master.

Another typical method for doing this is to remove the master from the field environment altogether and to include it in a single central computer. The master function exists as a task in the central computer. Data is transmitted between these pseudo-masters through shared memory (common) or other concepts.

The concept of field storage of the timing plans and the schedulers remains. A supervisory task in the central computer contains the database which indicates the pseudo-master to which a controller belongs. With this cross-reference data, the traffic engineer can define a traffic section with 'n' intersections in more than one master. The master (or channel or line) to which an intersection is assigned is totally transparent to the user.

Additionally, reports can be generated on the basis of traffic engineering sections and not be limited to channel assignments. Traffic responsive operation can be based on data collected across a series of pseudo-masters, not from intersections on a single channel, etc.

This "distributed system" concept, therefore, has the advantage, over UTCS, of being able to have up to 32 intersections per channel. It has the disadvantage of not having volume and occupancy data available for traffic responsible operation with a resolution of one second.

Distributed System Architecture

Figure D-6 illustrates the distributed system concept without field masters.

Note that this figure shows that the central computer is a bigger machine than required for closed loop type systems. It is not, generally, as large as required for a UTCS system. Because the central computer has more power than a PC, the design usually includes the ability for multi-users to operate with the system simultaneously. Usually, the user terminals are PCs which interface with the mainframe computer via a network, such as ETHERNET.

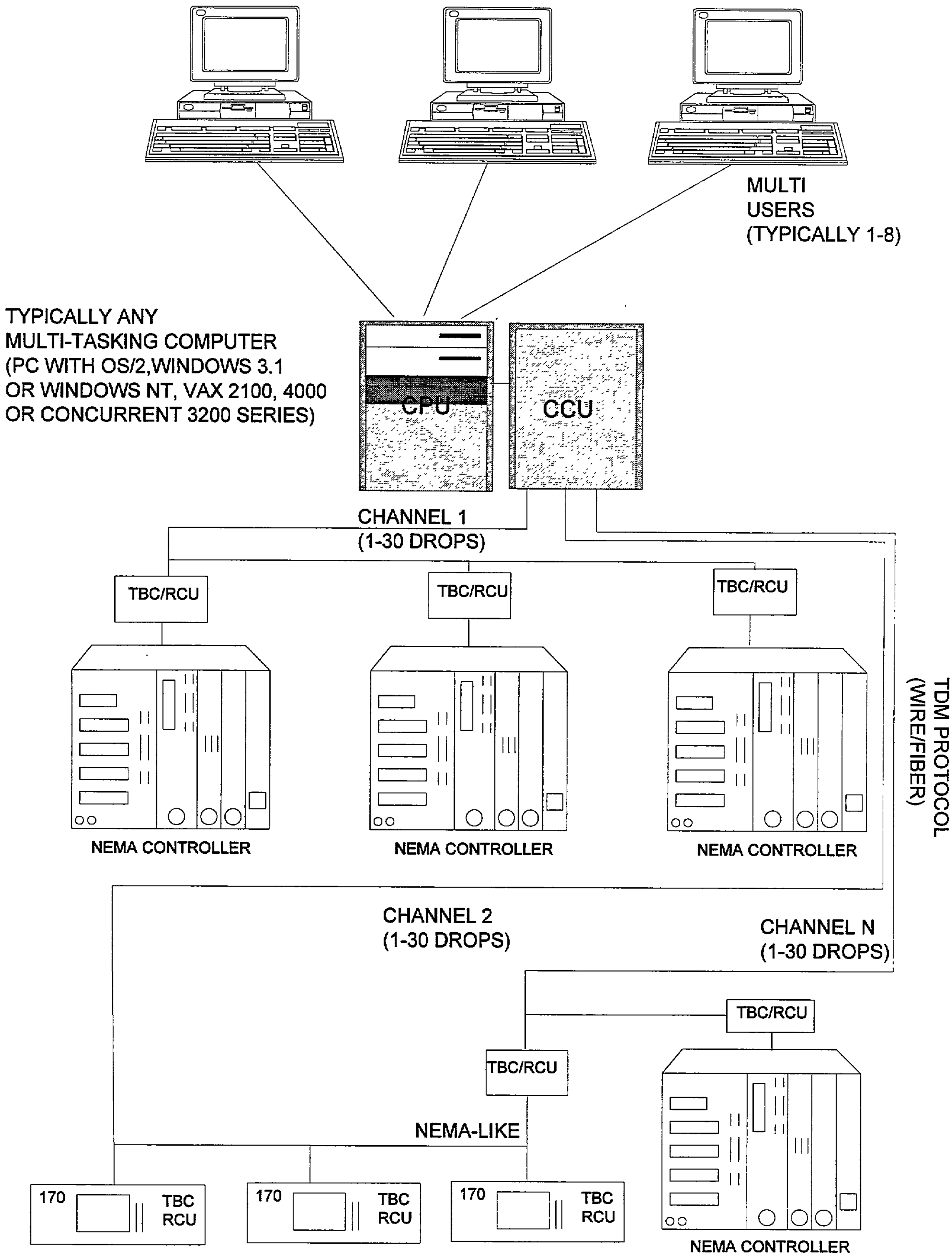
Note that in this illustration, the TBC/RCU assembly is shown external to the controller. This approach is very similar to the closed loop hybrid system described in the preceding section. This concept is not necessary. Some of the current distributed systems utilize this concept since the manufacturer does not make a controller.

Some of the current closed loop vendors have already migrated upward to the distributed system but their systems will have internal TBC/RCUs. This approach will, of course, have the same consequences of the closed loop system; the manufacturer will be sole-source for subsequent system expansion.

Other manufacturers who entered this field have maintained concept of having masters in the field. This requirement is, for the most part, necessary so that the user can readily migrate upward from a closed loop system to a distributed system without the cost of making major field wiring changes.

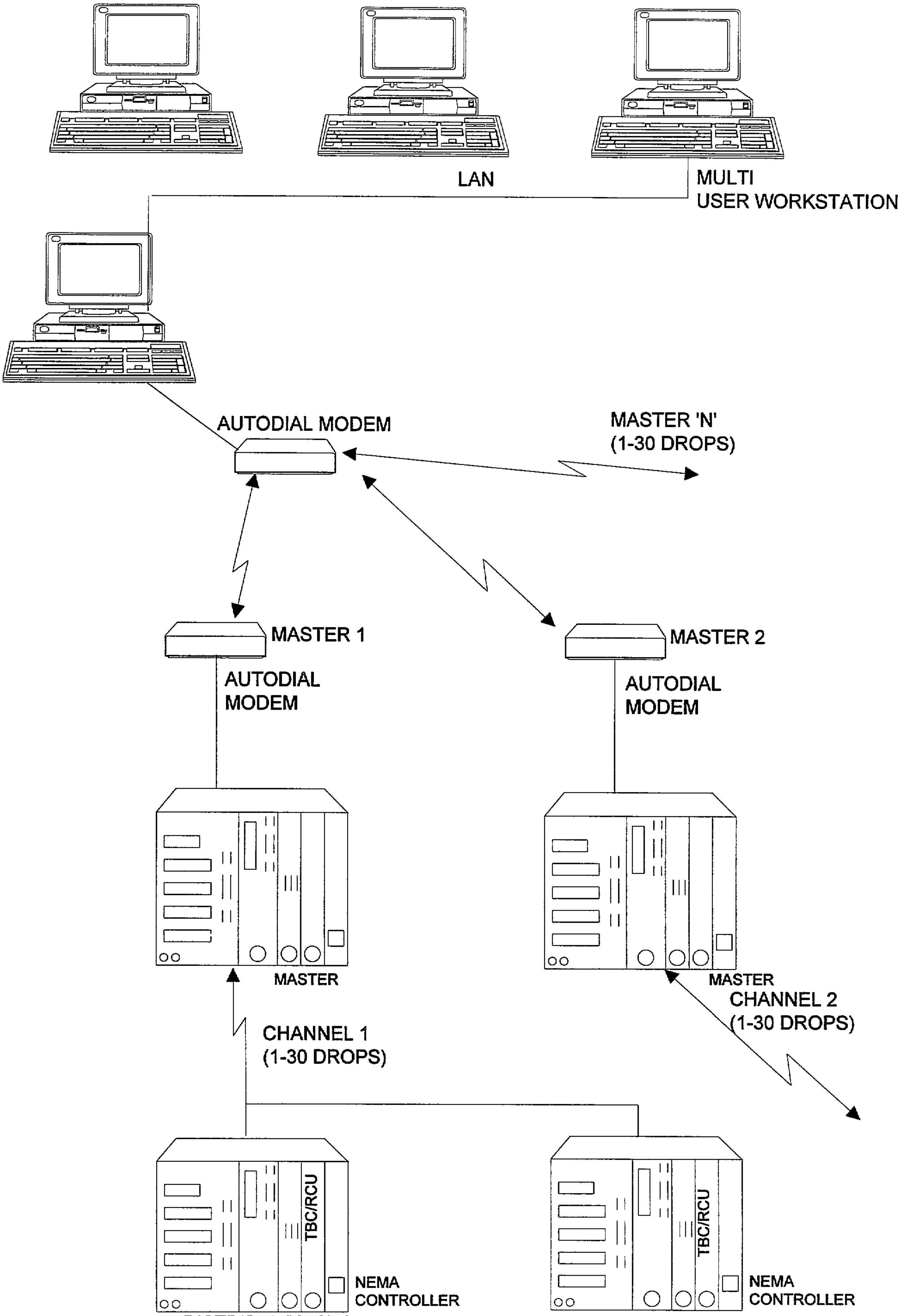
Figure D-7 illustrates the distributed system with field masters. In this configuration, the masters provide somewhat limited service when a traffic control section is defined across two masters. In this case, such functions as traffic responsive operation must be computed in the central computer, rather than in the master. The only known vendor of this type of

Figure D-6



DISTRIBUTED SYSTEM ARCHITECTURE, WITHOUT MASTERS
(any controller, no upload/download (except in 170 case))

Figure D-7



DISTRIBUTED SYSTEM ARCHITECTURE, WITH MASTERS system is PEEK/Transyt Corp. Its current name is the MATS system (Multi-Arterial Traffic System).

In either case, this approach maintains the closed loop architecture feature which has the once-per-second processing done on the street.

Because of this, up to 30 drops can be placed on single channel (as 1200 bits/second). With this sizing, and the capabilities of a mainframe computer, and with most of the one second processing on the street, this architecture allows for a much larger system than any of the other approaches.

Design Features

No Restriction to the Traffic Engineering Section Concept

The basic closed-loop system design was predicated on the fact that each closed loop system would correspond to one arterial. This design permits the user to assign any controller in his jurisdiction to any traffic control section without restriction. This applies to log and status reporting; to data averaging and manipulation; to timing plan scheduling by section; to traffic responsive operation by section, etc.

User-Interface and Database Management

The distributed system vendors have based their system concepts on closed loop system user-friendliness concepts. In general, the concepts are identical with the additional friendliness derived from being able to assign devices to traffic engineering sections. These systems use color graphics concepts for both monitoring and database management purposes. They can support multiple users. They can support massive amounts of background processing, etc.

Algorithms

Little has been done to improve the algorithms in these systems over those used in UTCS systems or in closed loop systems. One advantage that they do have is that since the traffic responsive algorithms are executed in the mainframe, its computing capacity provides a greater degree of latitude in the complexity that might be permitted in implementing future algorithms.

Advantages and Disadvantages of the Distributed System Approach

Disadvantages

- Field masters are removed from the streets. Autodial calling techniques are no longer required. To some, this may be an advantage. To others, the added cost of supplying channels (lines) from a central site to each intersection might be a burden.

- The communication system is proprietary, as in the closed loop system. A central cannot talk to any other vendor's coordination unit except for the PLUS system.

- Since all of the algorithms are stored and computed in the field unit (except for traffic responsive), any algorithm change requires changing out the PROM module in one or more units at all of the locations in the system.

- Since the time base must be maintained equal by all of the controllers in a section, this becomes a critical factor in the maintenance of coordination.

- Log data and sensor data can be lost if, for any reason, the central computer does not ask for it to be uploaded before it is overwritten by later data. The amount of data lost depends upon the size of the data buffer and how often it is uploaded.

- Volume and occupancy data are collected and stored with a resolution measured and averaged in terms of minutes, rather than seconds. This may not be a problem at the present time; it may be a problem later as new algorithms requiring high resolution data are developed.

- Complexity of on-street equipment is high. This does not appear to constitute a problem since modern solid state electronics operates very well in the hostile street environment.

- Without on-street masters, communication channels must be continuous from central to each street unit. Although leased telephone lines are not required, some users may find leasing costs more attractive than capital costs.

Advantages

- Traffic engineering sections can be addressed. This a major advantage over the closed loop type of system.

- Alarm data can be readily uploaded in real time. Even when the central computer is monitoring one channel, the multi-tasking capabilities of the mainframe permit it to be monitoring other channels as well. As a result, alarm data is available to the user almost as quickly as in a UTCS system. A maximum of 5 seconds delay between an alarm condition and its recognition at central is maintained.

- As with a closed loop system, communication costs can be reduced since communication bandwidths are lower. The amount of saving depends on the type of lines, the local tariff arrangement, etc.

- User interface capabilities are much enhanced, especially with respect to the older UTCS system approach
- Central computer costs are lower than for UTCS, but higher than for a closed loop system.
- No backup system is required. the basic control concept and the backup control concept are identical.
- Because field masters are removed from the street, on-street complexity is removed.
- There is no need for auto-dial modems or leased telephone lines to tie the central to the on-street network.

D. Data Processing

1. Detector Data Processing

The processing of the data collected from the vehicle detection system requires that a balance be maintained between locality of data availability, processing capability, communications circuit loading, and access to the data for analysis and presentation. Three options are typically considered:

- Transmit the data to a central location every second.
- Aggregate the data in the field for a specified time period (typically 20 set, 30 sec, or 1 minute), and transmit the aggregated data to the central location at the end of the collection interval.
- Aggregate the data over a collection interval (20 sec, 30 sec, or 1 minute), store this data in the field for an extended time period (up to several hours), and transmit it to the central location when required. The requirement for the data can be based upon an "event" occurring in the field, such as the detection of an incident; or when requested by the central system.

The first option requires relatively few bits to transmit vehicle counts because of the limited number of vehicles passing by a detector in one second. -Lane occupancy and vehicle speeds require about 10 bits per data item, in order to maintain accuracy. This combination of number of bits to transmit, and the one second transmission frequency places a heavy burden on the communications network (typically 1200 bps). It also requires a central computer system able to handle the data volumes and the data updates every second.

A second by second update is required when monitoring an arterial intersection controller, or an individual freeway monitoring computer. This monitoring is typically required for only a few such controllers simultaneously, so the overall system design need not provide. the capability for every location to communicate with the TOC every second.

Option two utilizes the power and processing capabilities of currently available microprocessors. As the processors that are deployed in field locations become more powerful and less expensive, distribution of the data processing is advantageous. This lessens the load on the communications network, and reduces the need for a larger central computer. The dynamics of traffic flow, and the rate of update of status maps and displays at the TOC establish the frequency of data transmission from the field devices. Operational experience has shown that updates every minute are not frequent enough, and updates every 10 seconds appear to be too frequent. This range has resulted in a 20 or 30 second communications time interval being utilized by several operational systems.

With this option, the field processor collects data for the selected time interval, and stores it in an intermediate data buffer until polled by the central computer at the TOC. There are numerous operational results, levels of service and summaries that can be calculated from the collected data. Since these calculations are equally simple to perform at either the field processor or the central computer, no advantage is gained by transmitting these derived values to the central system. They can be computed on an as-needed basis at the TOC (or other location) from "raw" field data "less expensively" that transmitting them. If they are needed at the field processor, for example by a technician reviewing the operation of field equipment, they can be calculated at that time in the field. This requires that the field processor have sufficient memory to store several hours (or days) worth of data. Computer memory in the megabyte (million byte) range is now very inexpensive, allowing this strategy to be implemented.

The data collected from an induction loop in a 20 second period can typically be represented with three bytes of data, and speed/length/classification counts obtained from a speed loop pair require less than six bytes of data. Thus, with six main-line lanes, one entrance ramp and one exit ramp being monitored, six speed pairs and four individual loops would be utilized. This results in about 48 bytes of data, plus overhead of about 20 bytes, being transmitted between the central computer and the field controller each 20 second period.

The case noted above where second by second monitoring of a controller is required must be included in the design of a periodic data collection/polling system. Since 20 second data. collection and second by second reporting are both equally important,. the communications system must be designed to permit 20 second data collection to be interleaved with second by second reporting. This interleaving must occur in a manner that does not exceed the delay requirements of either data stream, and fits within the available bandwidth of the communications channel.

The third option is useful when routine, periodic refreshing of status maps or data displays is not required. A data collection example would be the transmission of stored volume/occupancy data from the second loop of a speed loop pair only on as requested basis. Another example would be an incident detection algorithm running in the field microprocessor based upon variations in speed of individual vehicles - detail data that is lost when speeds are averaged over a 20 second period. Error reporting also falls into this category, since errors are infrequent events and need to be reported only when they occur.

The goal of most traffic monitoring and management systems is to reflect the “real-time” status of the roadway at the TOC, or other centralized location. This requires that data be transmitted from the field to the central computer on a regular basis. However, as noted in the examples above, there are categories of information that are infrequent (errors or detected incidents), or stored data that is needed only on a “demand” basis, or data that is available in the field processor but normally not used at the central computer (standard deviation of speed.) All of these situations require that the communications protocol and data formats be flexible enough to allow the system user to request or receive notification of this occasional data.

Computer Application

Tunnel Operations Monitoring And Control (TOMAC): A TOMAC IVHS system was recently implemented for the l-64 Hampton Roads tunnel complex. This control system has been applied at several facilities including the Elizabeth River Second and Third Downtown Tunnels, and the new four tube I-95 Fort McHenry Tunnel in Baltimore.

The system operates with automatic incident detection based on a modified California algorithm using absolute and relative occupancy. Detector communication is performed in one second increments, with a small degree of pre-processing, to convey accurate occupancies. When the software determines that a detected occupancy is likely to have been caused by an incident, the suspected incident is reported to the operator. An adjustable threshold of sensitivity is used. An excessively low threshold can result in a high false alarm rate, however an overly high threshold may result in missing a real incident. The occupancy threshold is automatically adjusted every four hours to compensate for various expected recurring traffic conditions including: AM & PM peaks, balanced and light traffic flow.

When a suspected incident is detected, the location of the suspected incident is identified and the operator is notified. The operator examines CCTV monitors to determine the nature and validity of the reported condition. The required emergency operation procedures are then manually entered. The Fort McHenry version of TOMAC is capable of entering emergency operation procedures without operator intervention, however the system is not operated in a fully automatic mode. The emergency operation procedures have been developed in a rudimentary expert system that controls variable message and lane use signs in the tunnel. The course of emergency action is dependent upon the current state of the tunnel and the location of the incident. The TOMAC System assists in changing all signs forward and behind the incident to appropriate states. Variable message signs (VMS) and changeable speed limit signs that have been pre-programmed are commanded for display under the appropriate circumstances.

The ability of TOMAC to detect possible incidents and direct an operator to a specific monitor to observe and verify a possible incident is beneficial in the operation of an automated incident detection and management system.

2. Incident Detection Algorithms

An incident is usually defined as any event that causes a temporary reduction in the capacity of a facility or roadway. Incidents may result from occurrences that physically block a portion of the active roadway or from occurrences entirely off the roadway that cause “rubbernecking” or “friction” effects. When a roadway is operating at a level below its capacity, an incident that reduces capacity, but leaves the roadway with enough capacity to handle the existing traffic produces few effects in the traffic system and will be difficult to detect by traditional means.

Various algorithms have been developed to perform automated incident detection. Different traffic parameters are measured and compared in a number of ways, each variation resulting in a new algorithm.

Traffic Parameters

The standard parameters used to quantify traffic are:

Occupancy - The percentage of a given time period that a vehicle covers a particular point on the roadway.

Speed - The average velocity of vehicles passing a point on the roadway during a given time period.

Volume - The number of vehicles passing a point on the roadway during a given time period.

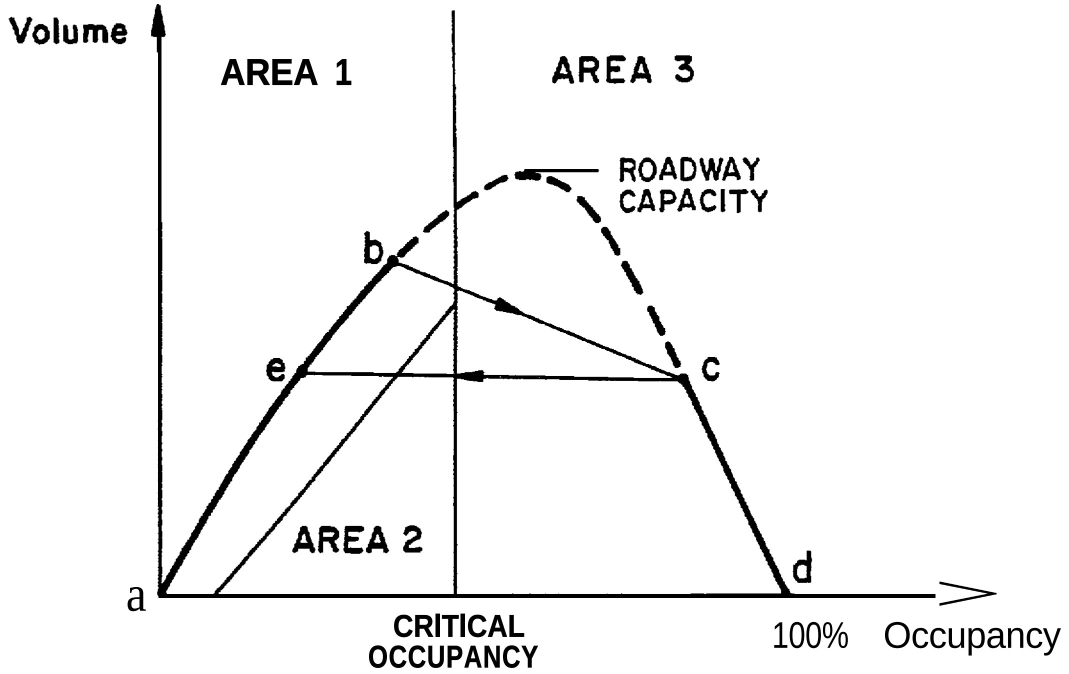
Comparisons of different types of time averages and new data as well as data from adjacent detector stations are the basis for incident detection.

Recurring congestion, due to operation of a roadway above its capacity, may be detected as an incident by some algorithms. A means of incident verification is needed to determine the cause of detected congestion. Incidents that occur on an already congested section of roadway are also difficult to detect, as this situation also fits the deformation of operation below capacity.

Incident detection algorithms measure and compare various parameters of the traffic stream to perform their function. As described in the literature, traffic tends to flow with a direct linear relationship between volume and occupancy (See Figure D-8, Line segment ab) under normal conditions. In congested operation, the relationship is shifted to restricted flow resulting in decreased volume and higher occupancy (See Figure D-8 Line segment cd. Point d represents standstill conditions.)

FIGURE D-8

FUNDAMENTAL VOLUME - OCCUPANCY RELATIONSHIP



AREA 1 = UNCONGESTED OPERATION

AREA 2 = LOW SPEED OPERATION

AREA3 = CONGESTED OPERATION

If an incident occurs at a time that the roadway is operating at point b, the volume will be reduced and occupancy will be increased to provide operation at the incident as shown by point c. Downstream of the incident, operation shifts from b to e.

Direction of Incident Conditions

When a queue develops from an incident, a shock wave travels upstream as additional vehicles are added to the queue and a metered wave travels downstream due to decreased volume and occupancy in the free flow after the restriction. The waves eventually reach detector stations where their effects can be sensed.

Detection of the metered wave that travels downstream at the highway flow rate may provide a rapid indicator of the occurrence of an incident. Detection of the shock wave, resulting from queue build-up traveling in the upstream direction provides further indication of an incident. Normal traffic flows that result in detection of parameters similar to the effects of both the metered and shock waves could be a result of normal bunching of traffic or "noise" in the flow causing false alarms.

The time taken by an algorithm for calculations to provide incident detection is usually not a major factor in detection time. The comparisons between algorithms usually rest on the time that it takes the parameters at the detection stations to reach values that the particular algorithm requires before declaring an incident.

California Algorithm

The California algorithm relies on the detection of three parameters or features between an upstream and downstream detector that are specific to incidents. These features must exceed all three specific thresholds:

OCCDF - spatial difference in occupancies >T1

OCCDRF - relative spatial difference in occupancies >T2

and DOCCTD - relative temporal difference in downstream occupancy >T3

Where OCCDF is the absolute (arithmetic) difference in occupancies, from data in the same time period, between upstream and downstream detectors. OCCDRF is a test to determine the relative size of the difference by dividing the absolute difference by the occupancy found at the upstream detector. DOCCTD is similar to OCCDRF except the test compares downstream occupancies at different times, time (0) and time (-5).

The modified California algorithm simply uses a different time period for comparison of DOCCTD, the times used are time (0) and time (-2). This results in a shorter interval used for the comparison test. A number of variations that improve some aspects of, or provide additional features for the basic California algorithm. Some of these variations allow the detection of incident termination, others provide less sensitivity to compression waves in the traffic stream, and others offer improved detection or have lower false alarm rates. A combination algorithm could be developed to provide the features desired at the project site.

Time Series Models

Another class of algorithms uses recent past occupancy history to model,. through time series, the near future values of occupancy parameters. When the projections differ by more than the threshold, an incident can be declared. Various statistical measurements of traffic parameters are used to detect incidents. The standard normal deviation algorithm considers the mean and standard deviation of occupancy over a period of about five minutes. An incident is declared when the measured value differs by more than one standard deviation from the mean. The double exponential smoothing algorithm employs absolute error between the observed and predicted value of volume and occupancy for one minute intervals. The ARIMA or autoregressive integrated moving average algorithm declares an incident when the observed occupancy is found to be outside the 95 percent confidence limit. The time series approach will detect congestion as well as incidents and does not offer any advantage regarding false alarms

McMaster Algorithm

This is a single station algorithm that operates on two-dimensional classification of flow and occupancy. The algorithm basically relies upon the determination of the roadway operating volume-occupancy region as shown in Figure D-8. A congestion flag is raised when operation is in area 2 or 3, or a slow highway speed is detected. When the flag is present for a specific number of consecutive periods, a potential incident is signaled. Since speed calculated from a single loop detector is unreliable due to a non-homogeneous traffic stream, most systems use paired loops to extract speed. The logic is claimed more efficient if data is collected from a lane with few or no trucks, as trucks tend to disrupt normal traffic flow. This algorithm tends to be successful at detection of congestion, however video confirmation is recommended to determine the cause of the detected congestion. Later developments of the algorithm apply comparison logic. Once congestion is detected, a check of adjacent station conditions is used to test for incident caused congestion.

Euocc

Developed for the British UK Transport and Road Research Laboratory, HIOCC seeks to identify slow moving and stopped vehicles. When several consecutive seconds of instantaneous occupancy are found to exceed a threshold, an incident is identified. Separation of incidents from other types of congestion is not performed by the algorithm.

Other Algorithms

There are numerous other algorithms that are either variations on standard algorithms or are experimental and are still in development. These include the Minnesota algorithm that utilizes data filtering and assumes that a large deviation in a system parameter must be caused by a malfunction. The Willsky algorithm that uses macroscopic modeling of traffic flow. The Cremer algorithm that models reduction of capacity at incident locations by considering an imaginary volume input at the incident location. Algorithm #7, as it has been referenced in the literature, is an adaptation of the California algorithm. The analysis of these and other possible algorithms will not be considered here.

Video Incident Detection

Several approaches to incident detection described in the literature are being applied as a result of video detection capabilities. Video detection offers the possibility of wide area detection from a camera location. One algorithm, called Speed Profile Incident Evaluation System (SPIES) employs several speed traps in each lane of traffic. The speed traps are positioned a few hundred feet apart to allow the system to analyze speed changes within view of the camera. Speed data is gathered and smoothed on the basis of volume resulting in samples about every 15 seconds. The samples are compared with 15 minute data in a historical database. An incident is detected by comparison of speeds measured at the upstream and downstream detectors, with expected speeds from the database and an alarm threshold.

Another video detection algorithm called Autoscope Incident Detection Algorithm (AIDA) uses the variation of traffic flow data with regard to both time and distance. A rapid breakdown such as a speed-drop or occupancy increase and speed thresholds are used to determine congestion levels.

A new type of incident detector is in development for Autoscope by Image Sensing Systems Inc. Although still in development, a January 1995 product announcement is anticipated. This incident detector is designed to reside within the Autoscope and can be configured on the unit's monitor in a fashion similar to the way standard detectors are configured. The incident detector will sense an unusual drop in speed coupled with an increase in occupancy. Comparisons are made with recent history using a dynamic threshold which automatically adjusts with the amount of traffic detected. Incident detection is expected to be available as an upgrade to older Autoscope models. Detected incidents will be reported on a serial communications port of the Autoscope using a non-proprietary multidrop protocol.

An evaluation of a traffic scene concentrating on the two dimensional data provided from a video camera is performed by another prototype system, Image Processing for Automatic Traffic Surveillance (IMPACTS). This system evaluates the spatial distribution, movement, and stops of traffic in the field of view. The roadway is divided into small areas called cells. The algorithm determines the magnitude of change for three variables in each cell. Spatial occupancy, weighted occupancy and lane state are tracked and relational changes in these variables are used to detect incident congestion. In a test near London, GB which ran for 170 hours, a total of 74 incidents were properly detected, two incidents were missed and four false alarms were logged. This system shows good promise for further development.

Conclusions

It appears that for the present if a high percentage of incidents are to be detected, a high false alarm rate will need to be tolerated. A false alarm rate of one percent will result from one false detection in one hundred tests. A technique is described in the literature which proposes a means for distinguishing between recurrent and incident congestion. This approach is good for single station algorithms but will not prevent false alarms. If data from the IMPACTS tests is reliable, systems of this type may provide true wide area detection. Since there is little experience with video detection and its ability to perform incident detection, the use of this technology may involve some degree of risk.

Detection algorithms cannot be expected to find every incident, nor can they be expected to perform without false alarms. Automatic detection of incidents provides the operator with a source of information about possible incidents on the roadway. The information obtained must be confirmed and dealt with no differently than any other source of information that may need to be questioned.

Several major factors surface from a study of incident detection algorithms:

- Single station algorithms detect congestion. A secondary means must be used to verify conditions and determine the cause of the congestion.

- A multistation algorithm will be dependent upon the continued operation of each detector. When one station fails, three comparisons cannot be performed. If detector stations are placed every quarter-mile, a half mile section would be lost with a single failure. It should be possible however, to bridge over the failed section and perform the comparisons with data from the adjacent detector stations.

- Incident detection requires that traffic parameters be checked for operation outside of certain thresholds. When data is found that exceeds some threshold by a large margin, an incident could possibly be reported without the need for additional testing. If the threshold is exceeded by only a small margin, further testing is justified. This type of magnitude testing might allow some incidents to be reported more quickly. Some algorithms consider overly large margin variations as probable hardware faults. Tests could be included in any developed algorithm to consider this possibility.

- Variable thresholds may be applicable for various levels of traffic and may be adjusted by time of day and day of week. For example, it may not be desirable or necessary to detect the recurring congestion at the start and end of each workday.

3. Traffic Signal Algorithm Alternatives

Introduction

Traffic control algorithms have lagged well behind the capability of the computing systems to deliver the timing plans that the algorithms develop. In the United States, particularly, there has been little effort made to develop new and more refined algorithms. Fortunately, both Europe and Australia have made substantial efforts to improve the algorithm field.

Algorithm Development Background

The UTCS system was developed starting in 1970. The algorithms employed in it (CIC, Traffic Responsive, data averaging, data smoothing, etc.) were all developed during the period 1970 to 1975.

In the infancy of traffic control systems, much thought and attention was given to the problem of how to move timing plan information from a central point to the field. In essence, that is the complete roll of a UTCS system. During that infancy, little attention was paid to how to develop the timing plans. The UTCS-Extended and UTCS-Enhanced programs had no capability for and scant attention had been paid to the problems of timing plan development. (The closest either came to a timing plan was in their attempt to change splits within the context of the existing timing plan (generated elsewhere). The FHWA spent considerable money in the development of a 2nd Generation system and the British spent considerable effort on the SCOOT system (3rd Generation), both which concentrated (with different approaches) on developing plans in real-time without Traffic Engineering assistance.

During the last ten years, while remarkable advances have been made in hardware and supporting software, the state-of-the-art in Traffic Control System software technology has remained relatively static. With the absence of the same kind of Federal funds that were once available and led to the UTCS-Extended and UTCS Enhanced software packages, there has been little impetus for the private sector to develop advanced software techniques.

To this time, neither UTCS nor any of the closed loop/distributed system vendors have made any change to those or original algorithms. In other words, traffic control capability has not been enhanced one iota by all of the new concepts. Only very small advances have been made above the original three-dial, three-offset capability offered by vendors in the 1950-1960 era.

The problems with today's traffic systems can not be resolved by new central system hardware or new control equipment on the street. UTCS offers the same capability to deliver timing plans to the street as to the close loop/distributed systems. The motoring public sees the same timing plans in either approach. The number of stops and delays causing driver frustration are the same with either system.

The future of traffic control systems will be in advanced algorithms. This will come about for two reasons:

- The traffic engineering community will soon realize the essential proof of the above and will start to demand better traffic control algorithms

- The rapid increase in the speed of computers will permit very complex algorithms to be developed which will run in real time.

A Review of Generations

1st Generation

The newer closed loop and distributed system approaches provide a new way of implementing timing plans; they do nothing to make it easier to develop the timing plans. These systems still rely on the Transyt-7F and Passer-II off-line programs (plus a lot of Traffic Engineering know-how) to make them work. These all use the "1st Generation" approach.

The timing plan is generated off line, usually using Transyt-7F or Passer-II, and is assigned to the traffic control system's database for implementation.

Generally speaking the current UTCS 1st generation philosophy is used throughout the industry. There are enhancements to it available. Some vendors call their systems something else other than UTCS, but they maintain the once-per-second communication with the field equipment and their traffic-responsive operation is the same as the WCS-Extended version.

Although the UTCS-Enhanced version was developed by the FHWA as an improvement to the original UTCS-Extended version, it had many problems which kept it from being used extensively. Recent improvements in machine speed may make WCS-Enhanced a viable alternative. Recently, the State of North Carolina had a successful implementation in the City of Raleigh.

2nd Generation

The technique which was employed in the 2nd Generation software version was to put Transyt-7F on-line. i.e., the data was collected by the system, fed to the Transyt-7F program, and, after considerable delay while the data was being processed, the resulting timing plans were developed and automatically output to the street.

At the time this work was being done, the machine speed capability was probably measured in terms of 0. 1 MIP. (The developmental machine was a Xerox Sigma 5, a long-obsolete machine.)

Because of the long delay in the time between data entering the system and the timing plan output from the system (15 to 20 minutes), the timing plan was usually not the correct one for the conditions that prevailed when it was brought in. The FHWA spent considerable effort at developing a “predictor”. The purpose of the predictor was to estimate what traffic flow would be like in 15 to 20 minutes, based on what it is now. The predictor study concluded that traffic was essentially unpredictable, but knowing the weather conditions improved the predictability greatly. (In the original UTCS development, the predictor was assigned to a second Xerox Sigma-5 while the 2nd Generation program operated in the first Sigma-5.)

The speed of modern and yet-to-be machines will make the predictor unnecessary. The time that it will take for the real time machine to compute plans using Transyt-7F will be so fast that there will essentially be no delay.

However, the 2nd Generation concept also requires a greater number of system sensors than does the 1st generation approach.

There is little general knowledge currently available concerning what is the most cost-effective number of timing plans to have on hand. A rule of thumb would indicate that it would not be very cost-effective to have more than 24 * 4 = 96/day since transition time would not allow the system to settle down any faster than that. However, the palette of timing plans from which this 96 could be selected could be huge. Certainly, the same 96 plans for a clear, dry day would not be drawn from the same group as for those required for a cloudy, wet day.

However, with only a limited number detectors available in the 1st Generation on concept, the data available for generating plans in a 2nd Generation system would not be adequate for the development of the infinite timing plan capability of that approach.

1.5 Generation

This can be thought of as a combination of the 1st and 2nd Generation schemes, In this system, the data collected by the traffic control system is automatically converted to the Transyt/Passer formats, the timing plans are developed and the Traffic Engineer can edit them at will. When he is satisfied with a timing plan, he can then send it to the system’s database automatically converted to the system’s format. (Data required by Transyt/Passer not automatically collected by the system are automatically retrieved from a historical database, prepared by the Traffic Engineer from whatever resources he has available.)

Supporting software for the 1 st Generation and 1.5 Generation approaches, (TRANSYT-7F and PASSER-II), are readily available in both PC and mainframe formats.

Both versions are fully supported by agencies of the Federal Government. 1.5 Gen versions will soon be available where the timing plans can be generated on either the mainframe or on a PC, depending on the users hardware resources.

3rd Generation

Third Generation techniques are best exemplified by the British SCOOT system, the related Australian SCAT system, the US OPAC system (in early development phase), the PRODYN system, and the UTOPIA system. Whereas the 1st and 2nd Generation systems require a "system" concept and imply that the cycle length of each signal in a section be the same (or a sub-multiple), these 3rd Generation systems compute force-offs (or advances) on a (roughly) four second basis without consideration of a "standard" cycle length or offset. These systems utilize one system sensor per approach (as opposed to 1 to 1.5 per intersection for lst/2nd Generation systems) and their optimization technique utilizes all of this data to minimize stops or delays in any way possible. It does not necessarily mean that a progression is the best way of arriving at these minimums (although, of course, it probably usually is.)

Third Generation and After

The British and Australians have each developed on-line optimization concepts with this approach. The British version (SCOOT) has been implemented in several British and Canadian cities and one US city with limited success. The Australians have implemented their (SCAT') system in several Australian and Pacific ring cities and one US City, also with limited success. The US developed OPAC system is currently under test as a single intersection device; it is has not yet been tested as a network control device. Other European systems (PRODYN, UTOPIA) are discussed later.

Both of these two systems operate in essentially the same way:

- Throw away the concept of cycle, split and offset

- Introduce a concept of "intelligent" actuation; i.e., make each controller act as if it was fully actuated but that the timing of the signal changes is based, not only on local conditions, but on conditions throughout the section in which the signal resides.

This is done by placing sensors in each approach (as in fully actuated operation), bringing the data back to a central site, processing the data through algorithms that optimize signal timing by minimizing stops and delays. The process is not unlike that used in the Transyt-7F program; i.e., many iterations are tried, each one being better than the last as determined by a "hill climbing" technique.

Of course, on an arterial, it is very likely that, for given conditions, a SCOOT timing plan will be the same as a PASSER timing plan i.e., a progression will be computed with an optimally wide bandwidth. However, again, realize that since cycle, split and offset are not fixed in a SCOOT system, it is difficult to directly compare the timing plans which result.

The important considerations of SCOOT, SCAT, PRODYN, etc., are:

- One detector per movement
- Data available at central with once-per-second timing resolution

From an operational point-of-view, what do these two considerations mean?

- System maintenance is a problem. Most jurisdictions have considerable difficulty keeping sensors operational. However, the British have argued, quite convincingly (based on experience in Glasgow and London), that it is cheaper to maintain sensors than it is to maintain timing plans.
- That the closed loop/distributed system approach is going in just the opposite direction than future algorithms will require; that is, once-per-second data resolution will be mandatory.

Suppose SCOOT or SCAT or similar approaches are not found acceptable. In current timing plan development techniques, saturation flow and turning movements are essential inputs to Transyt-7F and to Passer-11. Current traffic systems (UTCS or distributed do not go far enough to collect all of the data needed. Turning movements can be obtained if sensors are included on each approach. Saturation flows can be determined if detectors (of the right type) are included on each approach. Detector technology which currently exists can permit the collection of both turning movements and saturation flow rates even when the movement loops are of the long type (26 feet, 40 feet, etc.). This current technology can permit the collection of this type of data for introduction directly into PASSER or Transyt-7F)

It is quite likely that an approach of this type will be the bridge between the current off-line timing plan development systems (which use inadequate data) and the new methods as exemplified by SCOOT and SCAT (where timing is developed on-line based on massive amounts of data).

In any case, the requirement will be for data collection with once-per-second resolution, not for once-per-cycle, once-per-30-seconds or once-per-5-minute resolutions.

Current Algorithms State-of-the-Art

The UTCS Method

Many of the newer distributed/closed loop systems have adopted the UTCS method of traffic responsive timing plan selection. This method compares the “loss factor” (a fairly arbitrarily chosen MOE) of different pre-stored timing plans (including the one currently in operation). If the loss factor for any one of the stored timing plans is sufficiently lower than the current timing plan's loss factor, the new plan is put into effect. Generally speaking, the vendors provide for 4 or 5 traffic responsive candidate timing plans per time of day.

Other Methods

A few of the newer closed loop vendors have developed a traffic responsive strategy which selects the cycle length, based on some form of theory, they select the split based on the relative demand along-street vs. cross-street; they select the offset based on the relative demand in either direction on the main arterial.

Although these approaches have not been well evaluated by external sources, or by simulation, users of these systems seem to be satisfied that they work reasonably well. All of these approaches are designed for arterial applications; extension of the approach to networks is questionable.

Review of the OPAC Method

OPAC Features

Optimization Policies for Adaptive Control (OPAC) is a computational strategy for truly realtime demand responsive traffic signal control. As formulated and described by Dr. Nathan Gartner, it has the following features:

- It provides performance results that approach the theoretical optimum.
- It required on-line data that can be readily obtained from upstream link detectors.
- It is suitable for implementation on existing microprocessors,
- It forms a building block for decentralized control in a network.

Operation

The local OPAC controller operates on the dynamic programming principle. Future traffic stream arrivals are measured and projected on all approaches and an on-line optimization algorithm dynamically selects the best switching strategy. Signal control is optimized in terms of switching points rather than the conventional phase lengths. Thus, the conventional notions of 'cycle time', 'splits' or 'offsets' do not exist. They can only be deduced from the calculation of the switching points and they are continuously varying in response to the traffic demands.

The optimization procedure that computes signal switching times seeks to minimize a specified Performance Index (PI) over some specified duration called a "horizon. To relinquish the extensive requirements for both the calculation and prediction, the actual optimization algorithm uses a Rolling Horizon concept to implement the Dynamic Programming principle. That is, the horizon "rolls" as time passes and new detector information becomes available. For each such roll, the optimization procedure is repeated to calculate the timing of switching points to minimize a calculated Performance Index. There may be several signal switching points computed within one optimization calculation (over the current horizon). Flow arrivals are measured upstream by detectors and a "dynamic flow profile" is created for the entire horizon. The profile consists of a "head" portion and a "tail" portion. The head represents actual arriving vehicle detection and its duration related to the travel time from upstream detectors to the stopline. The roll period may be less than or equal to the head period. The tail has smoothed flow values. Thus, the tail profile provides a means for interaction among adjacent controllers (or, intersections). The horizon length (head plus tail) is on the order of magnitude of a typical cycle time in conventional, fixed-cycle systems; i.e., 60 to 100 seconds.

The optimization procedure includes an estimate of queue length, delay and occupancy (all smoothed estimates) which contribute to the PI. An Optimal Constrained Search (OCS) method is applied to determine the minimum PI and the associated switching times projected over the horizon. If a switching time occurs within the head period, it is implemented. The 'smart' signal can be implemented by a microprocessor-based controller at each intersection. It is envisioned that for flexible system operation, an upper level control unit will be required to provide coordinative communications with the local controllers, in order to achieve any of the desired strategy objectives. The upper level control unit, however, may be located at any of the local control units. The coordinative level unit may also conduct a variety of other functions such as on-line estimation of network performance measures, projections of both upstream and downstream flow patterns, etc.

Detector and Communication Requirements

At the present time, OPAC operates with conventional loop detectors located on all upstream ends of the signal approaches, the same as for SCOOT or PRODYN. The outputs of those detectors are linked to the downstream local controllers. In a multi-level hierarchy, these outputs will also be communicated to the upper level unit.

In the future, it is conceived that more advanced surveillance may be used such as wide-area video detectors, that will offer additional capabilities for the system operation.

Review of the SCOOT System

Introduction

SCOOT is a fully adaptive traffic control system which uses data from vehicle detectors to optimize traffic signal settings so as to reduce vehicle delays and stops. It was developed in the United Kingdom by the Transport and Road Research laboratory together from three UK signal companies .

Operation of the SCOOT System

The three key principles of SCOOT operation are:

i) Measurements of Cyclic Flow Profiles (CFP).

SCOOT, in common with the well known TRANSYT method is based on the use of "cyclic flow profiles". A CFP is a measure of the average one-way flow of vehicles past any chosen point on the road during each part of the cycle time of the upstream signal. The average flows in each part of the cycle can be taken over many cycle times (e.g., a I hour peak as in TRANSYT) or updated every four seconds (as in SCOOT). The cycle time is divided into short time steps which are fixed at four seconds in SCOOT. Thus, a CFP records platoons of vehicles as successive steps within the cycle when flows are high.

In TRANSYT, the shape of the CFP has to be calculated for each one-way flow along all streets in the area. The calculation is made in an upstream-downstream direction. The accuracy of the calculation depends on the accuracy of the data on average flows, saturation flows, cruise times and so on, that the traffic engineer has to provide as input to the TRANSYT program. It is not difficult to collect these data, but it is time-consuming. SCOOT bypasses these processes and achieves accuracy and immediacy by measure the CFP in real-time, but of course, vehicle sensors have to be installed and maintained to provide the data.

ii) Update of On-line Model and Queue Estimation.

Once a CFP is known the computer can calculate when vehicles will reach the downstream signals and hence both the size of the queue and how long it takes to clear, can be calculated and the effects of alterations in the signal timings predicted. TRANSYT and SCOOT carry out these calculations in a similar manner. Both methods assume that traffic platoons travel at a known cruising speed with some dispersion, and that queues discharge during the green time at a "saturation" flow rate that is known and constant for each signal stopline. The growth and clearance of a queue can thus be calculated. In SCOOT, these estimates are updated every four seconds and constitute the "on-line traffic model" that is used in real time by the signal optimizer. The traffic data for the CFP's are collected, usually every second, from inductive loop sensors located well upstream of signal stoplines, installation costs are reduced and the earliest possible direct prediction is obtained of arrivals at 'the downstream stopline. Further, the sensor can detect "gridlock", which may occur if the queue extends back into the upstream intersection. The SCOOT optimizer takes special action when vehicles queue over the sensor.

iii) Incremental Optimization

The third key principle is that the coordination plan should be able to respond to new traffic situations in a series of frequent, but small, increments. This is necessary because research has shown that it is very difficult to predict traffic flows in the next few minutes, hence any “fixed” coordination plan may be out of date before it is calculated or inappropriate after it is implemented (and implementation is likely to cause extra delay during the transition from the old timings to the new).

SCOOT uses an “elastic” coordination plan that can be stretched or shrunk to match the latest situation recorded by the CFP’s. This is achieved by optimizing the splits, offsets and cycle time. A few seconds before every phase change, the SCOOT split optimizer calculates whether it is better to advance or retard the scheduled change by up to four seconds, or to leave it unaltered. Then, once a cycle, the offset optimizer assesses whether the delays and stops on streets around each intersect on can be reduced by altering the offset to be four seconds earlier or later. Favorable cycle time of a group of intersections may be incremented up or down by a few seconds every few minutes. SCOOT will not, however, make any change to the order in which stages are run.

So SCOOT makes a large number of small optimization decisions, typically over 10,000 per hour in a network of 100 intersections. A few decisions may be wrong, but this is unimportant provided the large majority are correct.

Continuous monitoring of detectors takes place in order to identify if any are faulty. This monitoring will typically identify a detector as suspect faulty if it remains ‘off for more than six minutes or ‘on’ for more than three minutes. This may occur as a result of hardware failure, or alternatively as the result of parking or some other obstruction over or near detectors. If a suspect detector begins to register vehicles again, its suspect faulty status will be canceled, however, after approximately 30 minutes the detector will be identified as faulty and operator intervention is necessary to change its status. While a detector is faulty or suspect faulty, SCOOT will operate from fixed timings for the links concerned. It is recommended, for good operation of SCOOT, that at least 90 percent of detectors should be working and maintenance plans for detectors should be aimed at achieving this rate.

Control Strategy

Contiguous groups of intersections are placed into regions which will have a common cycle time. Several regions will be combined into a cell and controlled by a single computer. the number of intersections in a cell depends on the power of the computer used.

Detector Requirements

The detectors used by SCOOT are inductive loop detectors installed at the upstream end of each of each link in the network, typically at about 15 meters past the intersection. These may cover a maximum of two lanes per detector and are generally two meters in the direction of travel. Where there are a significant number of links and sources of traffic along a link or other physical problems, it may be necessary to locate a detector further downstream. This usually occurs, for example, for a right turning lane which is treated as a separate link. The detector is sampled every quarter second and data is transmitted every second.

Range of Operation

SCOOT has been operational in the UK since 1980 in Coventry. There are now around forty implementations within the UK, with the largest controlling the central part of London and other parts of Greater London. There are also systems in Bejing, Hong Kong, Santiago, Durban and Port Elizabeth. Further systems are proposed in Madrid, Cyprus and Nijmegen (Netherlands). There are four systems in North America; these are Toronto, Red Deer and Halifax (Canada) and Oxnard, California in the USA.

Theoretical and Practical Soundness of Ideas

There are a number of basic philosophies which led to the development of SCOOT. One of these was to provide a fast response to changes in traffic conditions to enable SCOOT to respond to variations in traffic demand on a cycle-by-cycle basis. Hence an optimization of the stage split is performed before each stage change. Offset optimization is performed once each cycle and optimization of cycle length occurs every five minutes or every two and a half minutes during times of rapid change. Optimization takes place to minimize delays and stops for vehicles. SCOOT responds rapidly to changes in traffic, but not so rapidly that it is unstable and produces large fluctuations in control patterns as a result of temporary changes in traffic patterns.

SCOOT decisions are based on current detector data, using the traffic model's short term predictions of queuing patterns. There is no attempt within SCOOT to model long term, historic changes in traffic patterns. The underlying signal plan is changed slowly as SCOOT responds quickly to cycle variations in traffic.

SCOOT's traffic model utilizes cyclic flow profiles which are updated based on measures of demand derived from detectors. These cyclic profiles are used in a manner similar to the off-line TRANSYT method of calculating fixed time plans, but updated every four seconds. Queues are estimated by assuming that vehicles travel at a fixed cruise speed (with some added dispersion) from the detector to the stopline and when signals are red these vehicles are queued, when green they discharge over the stopline at a fixed rate. The upstream location of SCOOT detectors allows for the detection of exit-blocking at an intersection and the SCOOT model will continue to queue traffic at the intersection when this occurs.

SCOOT is in use worldwide and has been shown to give significant benefits over fixed time operation. The technique has a sound theoretical background and has been proved to operate well in practice.

Graphics Capability

The specific graphic displays available vary depending on the system vendor, however, a number of basic types of display are common. TRL use their own set of graphical display programs which include intersection displays showing the intersection layout and real-time information on signal settings and queues for each approach. Another common display is known as a VEGA display, and this shows graphs of demand and queues for each approach. Graphs showing how offsets are chosen to minimize delay on each approach are also available. In addition an overview of a region, using color coding to display congestion is available. This information comes directly from the SCOOT model in real-time.

System Evaluation

The effectiveness of the SCOOT strategy has been assessed by major trials in five cities as shown in Table D-3. The trials in Glasgow and Coventry were conducted by TRL and those in Worcester, Southampton and London by consultants, a university, and the local traffic authority, respectively. In most cases, comparisons were made against a good standard of up to date fixed time plans, usually produced by TRANSYT. The following table shows that the largest benefits are achieved in comparison with isolated vehicle actuation but, of course, part or this benefit could be achieved by a good fixed time system. The relative effectiveness of SCOOT varies by area and time of day, but overall it is concluded that SCOOT achieved an average saving in delay of about 12 percent compared with good fixed time plans. Since SCOOT does not “age” in the way typical of fixed time plans, it follows that SCOOT should achieve savings in many practical situations of 20 percent or more depending on the quality and age of the previous fixed time plan and on the rapidity with which flows change.

On the basis of the surveys and subsequent experience, SCOOT is likely to be of most benefit where vehicular flows are heavy, complex and vary unpredictably.

TABLE D-3

PERCENT REDUCTION IN DELAY FROM THE USE OF SCOOT

Location	Previous Control	AM Peak	Off Peak	PM Peak
Glasgow	Fixed-time	-2	14	10
Coventry	Fixed-time			
Foleshill Rd		23	33	22
Spon End		8	0	4
Worcester	Fixed-time	11	7	20
	Isolated V-A	32	15	23
Southampton	Isolated V-A	39	1	48
London	Fixed-time	(Average 8% less travel time)		

"Fixed time" refers to "up-to-date" TMNSYT.-generated timing plans.

Results significant at the 95% confidence level.

"Isolated V-A" refers to uncoordinated full actuated operation.

System Setup

The SCOOT system requires no calculation of initial fixed time plans; previously used plans may in fact be used when a system is set up, however, there is no necessity for this. Fixed time plans can be derived from SCOOT for use as backups during detector failure for example at the signal stopline for each link. The information required to set up the SCOOT network is therefore primarily network details (such as which signal stages control which links), cruise times from detector to stopline and the rate of discharge of vehicles over the stopline (measured on street as the system is brought on-line). There is no necessity to specify critical links or intersections.

Ease of Access to Setup Experience and Information

Training courses, open to anyone, are available annually at TRL in the UK, and by arrangement in other countries. Due to the large number of SCOOT installations, expertise in setting up SCOOT systems is also widely available. Assistance may be sought from the traffic signal companies which supply SCOOT and from a number of UK Traffic Consultant who have experience setting up SCOOT systems.

Advanced Traffic Management System Features

Emergency vehicle green waves can be used on predefined routs in the network. These are usually used by ambulances or fire-engines. These are initiated by operator intervention and may either be implemented for a pre-defined time or until canceled by an operator. There are two types of implementation. The first is a 'fixed time' green wave facility where all the signals on the defined route run on a green wave plan throughout the duration of this operation. The second, and more recent, solution is a 'rolling' green wave plan in which only one intersection along the route is affected at any time and the plans are calculated to allow the vehicle to arrive at each intersection during the operation of the green wave plan.

Public transport priority features within SCOOT are at present limited and confined to favoring links with high bus flows using existing facilities within SCOOT such as split weighting and fixed and biased offsets. These are described in more detail below. Research is continuing within the DRIVE II project, PROMPT, into active bus priority measures.

Bicycle operation has been developed for SCOOT, using different shaped bicycle detectors, the data is processed differently and then can be incorporated into the usual optimizing process. There is also scope for priority weighting for bicycle links using normal weighting and biasing features within SCOOT. A system using the bicycle logic has been installed in Beijing, China.

Gating is a technique designed to limit the flow of traffic into a particularly sensitive area. The gating logic allows one or more links to be identified as bottleneck links where problems are known to occur. Associated gate links are identified where it is less critical if queues build up. Under normal conditions no gating action is taken, but when saturation on the bottleneck link reaches a defined limit then the optimizers will being to reduce green time on the gated links in addition to its normal optimization. An alternative operation of gating is to specify links downstream of the bottleneck which will receive increased greens when critical saturation is reached on the bottleneck link.

Route priority is permitted using fixed and biased offsets and split weightings. Without such measures, SCOOT will consider equally the delay to vehicles on any approach. For traffic management purposes there may be a requirement to favor particular approaches and to fix the offsets along a main road. The preferred solution, to avoid excessive delays on side roads is to use biased offsets and split weightings which will allow SCOOT to perform some optimization, but with regard to the desired management strategy.

On-line saturation occupancy measurement has been included in SCOOT. The saturation occupancy of a link is normally determined manually in the field when a SCOOT system is set up. The Saturation Occupancy/Flow Technique (SOFT) has been provided in order to reduce the amount of time required to set up a SCOOT system and to automatically vary saturation occupancy in response to changes in real-time due to factors such as on-street parking.

SOFT is not suitable for use on all SCOOT links, but a trial in London found that it was appropriate for some 60 percent of links,

Traffic management information is available from SCOOT's on-line traffic model which predicts queues and delays. Information from this model is available to traffic managers in the form of text messages which output information from the model, or in a graphical forms. A database facility, called ASTRID, has also been developed by TRL and Southampton University, which allows historic information from the SCOOT model to be stored and analyzed to produce graphs and tables of typical flows, delays and congestion throughout the day, and long-term trends for the same types of data.

Proprietary Rights

Proprietary rights to the SCOOT system are held by the Government of the United Kingdom and three UK Traffic System Companies; namely, GEC Traffic Automation Ltd, Peek Traffic Ltd and Siemens Plessey Controls Ltd.

Source Code Language

SCOOT is written in CORAL which was the most suitable programming language for real-time applications at the time of its early development.

Software Portability

SCOOT is portable onto difficult computers, providing CORAL compiler is available for the machine in question. From Spring '93 it was also possible to use a C version of SCOOT.

Availability of System in PC Compatible version

PC versions of SCOOT are already available.

Review of the SCATS System

Overview

SCATS is a centralized traffic adaptive system developed by the Roads and Traffic Authority of New South Wales, Australia. Details of the system are described in the sections which follow.

Operation of the SCATS System

The SCATS system controls signals in groups, known as sub-systems, where the critical intersection for each subsystem is specified by the traffic engineer. Sub-systems are grouped together and a regional computer can control signals at up to 10 intersections. Systems can expand by the addition of extra regional computers which control traffic in their own area, but a central monitoring computer usually controls data input and traffic monitoring to the different regional computers.

SCATS uses two types of detectors, known as strategic and tactical. SCATS collects flow and occupancy data from these detectors and calculates the degree of saturation for each link where strategic detectors are installed.

Cycle times are optimized each cycle. The split may vary each cycle to maintain equal saturation, based on the average degree of saturation on approaches over the last three cycles. To avoid oscillating offsets, the optimum offset is calculated each cycle, but only implemented when at least three out of the previous file cycles have suggested a change to that offset. The sequence of stages may vary and where there is no demand, stages may be omitted.

Each local microprocessor operated controller is connected by a dedicated voice line telephone link to the regional computer with a communication rate of 300 bps. The communications hardware and software is integrated within the controller. Each controller has 24 detector inputs, 8 pedestrian push button inputs and outputs for up to 16 vehicle signal groups and 8 pedestrian signal groups.

The central monitoring function can be performed in one regional computer. However, where regional computers are distributed, communications between regional computers and a central monitoring computer is carried out using modems (at either 1200 or 4800 bps). The regional computers are sufficiently robust to be located at outstations for reduction of cabling costs, although they are not suitable for roadside installation.

In the event a regional computer fails or there is a loss of communications links, the local controllers will revert to an autonomous form of control. The nature of this is described in more detail in the following section.

Control Strategy

SCATS uses two levels of control, known as ‘strategic’ and ‘tactical’. Strategic control is carried out by the regional computer and tactical control by the local controllers. Both levels together control the signal cycle time, split and offset. Strategic control is responsible for calculating signal timings based on typical traffic conditions, and tactical control optimizes settings at individual intersections, within the restrictions imposed by strategic control.

Strategic control uses flow and occupancy data from detectors to determine optimum settings for a ‘sub-system’ of up to (about) ten signals. For each sub-system, a critical intersection must be specified, which will have priority when cycle times and offsets are calculated. All intersections in a sub-system will run on a common cycle time, and settings for other links will either be defined as constant or be calculated to be compatible with the settings for the critical intersection.

Tactical control allows stages to be terminated early where less than average demand exists or to be omitted when no demand is present. The exception to this is that one stage (usually for the main road) for each intersection is specified which cannot be omitted or end early. This stage will use any extra time available in the over-all cycle time (was set by strategic control) which is not used by the other stages.

Detector Requirements

SCATS uses inductive loop detectors. These are located at, or near, the stopline. Strategic links are identified by traffic engineers and ‘strategic’ detectors, normally 4.5 meters long, are installed in all lanes. Occasionally additional strategic detectors are installed upstream in order to detect queuing. As described previously, the data from strategic detectors is used for regional control decisions. At the local control level, ‘tactical’ detectors are used which are normally of the same type as strategic detectors, but may be located upstream of the stopline.

This allocation means that detectors are not necessarily installed for every link in a network, also, for some less significant links only one detector may be used for two or more lanes. Strategic detectors are required on all links approaching major intersections and on any links which are immediately upstream of a major intersection. Tactical detectors are used, in general, for more minor approaches to identify turning movements. Hence, most links are likely to require either tactical or strategic detection. All detectors are capable of performing either function.

Range of Operation

SCATS has been used in Sydney, Australia since about 1975 and has a user base of 26 systems in Australia and New Zealand and further systems in Shanghai, Shenyang (China), Singapore, Sandakan (Malaysia), Rauia Lumpur and Dublin (Ireland). The only North American implementation is in Oakland County, California.

Theoretical and Practical Soundness

The principal philosophy of the control strategy is to minimize vehicle stops and delays and to maximize throughput when the network is operating near saturation. It utilizes microprocessor controls to supplement central control with a distributed control element.

The system is filly adaptive except in respect to those aspects which must be defined by the traffic engineer at the outset. The primary requirement in this respect is the identification of critical intersections within sub-systems. This may be easily done for existing operational systems under typical conditions, however, the results may be less satisfactory for new road networks or ones in which conditions are different at various times. This limitation can be overcome somewhat by using smaller sub-systems or different groupings at different times of day. However, this approach will affect cycle time and offset coordination.

SCATS has been installed in many locations and is fully viable in practical operation. It has been suggested that this success also relies on the skill of traffic engineers who set up the system.

Graphics Capabilities

A number of graphical facilities are available for use on PC compatible machines used as terminals by operators. These allow for displays at the individual intersection level, showing intersection layout, and real-time display of detector operation and signal settings. At the sub-system level, the layout is shown with graphs to represent flow and occupancy which is measured by the strategic detectors in the sub-system. Regional displays show color-coded maps with a real-time indicator of flows in the region. There is also a whole system display which gives an overview of traffic conditions in each region and details of the six regions with highest flows.

System Evaluation

A survey carried out in Paramatta in 198 1 by the Australian Road Research Board showed no significant reduction in travel times compared with operation using TRANSYT, however there was a large reduction in the number of stops, some 9 percent in the central area and 25 percent on arterial roads. Other studies have indicated improvements in travel times but compared to the original systems which were of unknown efficiency.

System Setup

In addition to standard parameters defining the network, each local controller in a SCATS system requires a timing Plan. This procedure is similar to collecting data for calculating TRANSYT plans and can be time consuming. Programs exist as part of the management computer software to reduce the effort of preparing these plans and other controller data which is required.

Ease of Access to Setup Experience and Information

Training courses are run for SCATS, and the large number of systems in operation means that it might be possible to gain knowledge from existing users.

Advanced Traffic Management System Features

Public transport priority can either be passive, by means of weighting system performance to favor public transport routes (by time of day if required), or active using detection of a public transport vehicle. Using the latter option, the facilities exist to switch to a particular stage on detection of a vehicle, or to retain a green stage if appropriate. An automatic adjustment of split then follows the changes made. SCATS has also been used to implement tram priority.

Emergency vehicle priority operates in the same manner as active public transport priority, as described above.

Bicycle operation has been developed for SCATS and is used in the system installed in Shenyang, China. This uses a smaller detector loop than for motor vehicles, but the standard software can be used for signal control.

Dynamic advisory speed signs have also been tested with SCATS.

Information is provided for traffic management purposes using the messages which are sent from controller to regional computer and can also be displayed to the operator-in real-time. No storage facilities exist for this data. The management function software also provides databases for storing count data (either form SCATS detector or other sources) and for storing travel time data from specially equipped vehicles, which is collected through the SCATS system.

Proprietary Rights

Proprietary rights are held by the Roads and Traffic Authority, New South Wales, Australia. All users in Australia and New Zealand have access to SCATS and the Australian signal companies Philips Traffic Systems and AWA Ltd. have an arrangement with RTA to market SCATS in other countries.

Source Code Language

The SCATS Regional Control computer is programmed in DEC PDP-11 Macro Assembler.

Software Portability

The SCATS software required a Digital PDP-11 series computer with 22 bit bus. Software required at least I Mb of memory, together with a hard disk and disk or tape backup device. A PDP-1 1/3 is sufficient for a region of up to 40 intersections or an 1 l/93 for up to 128 intersections. A synchronous multiplexer is required for terminals for every eight regional computers.

The Roads and Traffic Authority states (Lowrie 1992) that:

> "at this point in time there is no other combination of computer and operating system that even approaches the PDP- 11 in suitability...the use of PC, VAX or "UNIX boxes" is suggested from time to time but all are inappropriate due to lack of I/O capability, cost, wrong type of operating system or lack of development and guarantee of long term availability. "

Availability of System in PC Compatible Version

SCATS is not available in a PC version.

Review of the UTOPIA System

Introduction

UTOPIA was developed in Italy with the objective of improving private and public transport efficiency. Characteristics of the system are described in the sections which follow.

Operation of the UTOPIA System

The UTOPIA system places strong emphasis on decentralization of optimization, with signal controllers playing a significant role in the strategies. Local controllers determine signal settings, priority for public transport vehicles and adjacent controller coordination. The central computer supplies any overall constraints, or area-wide strategies. UTOPIA distinguishes between these two levels and uses open loop feedback optimization techniques to supply control at the local and area levels.

Detectors provide flow and occupancy data. Most provide incoming flow for downstream intersections and outgoing flow for upstream intersections. Additional queue detectors are installed at locations which are determined to be critical. This data is processed by intersection control units which communicate with the traffic signal controller, the adjacent intersection controllers and the area level controllers.

Control Strategy

The UTOPIA system may be described as a hierarchical-distributed system, and the system functions are divided between two levels of control: the area level (central control computer) and the local level (intersection controller).

Area control is the higher level at which an "observer" uses a macroscopic model of private traffic and aggregate traffic conditions, to predict traffic flows for private traffic. This prediction is based on past information (contained in a continuously updated database) and on new information coming from the lower level. The outputs of the area control are desired speeds and flows on the network expressed in a "reference plan" with constraints and weights to be assigned to the individual components of the local cost functions.

The local controllers (SPOT units) act on their own microscopic model of the local area considering information coming from the higher level. At every decision instant (a few seconds) their objective is to minimize a cost function over an interval on the order of a few minutes. The cost function takes into account both delays at the controlled intersection (with weights dynamically assigned by the area controller) and the correspondence of local decision and policies with reference decisions provided by the area controller. Furthermore, a correct integration of private and public traffic control is required. Therefore, the local controller is able to accept forecasts of public transport vehicle arrival times and to adjust offsets so that the public vehicle will arrive at the intersection during the green time.

Detector Requirements

The detectors are inductive loops which are sampled every eighteenth of a second. They are located just downstream of the previous intersection and are connected to the traffic signal controller. Experiments have been conducted using different types of detectors for the queue detectors, such as cameras. The central system and all intersection controllers have cable links for communication.

Range of Operation

UTOPIA was first used in 1985 in Turin. This is the only current operational system, but there are plans to implement UTOPIA in Gothenburg and Salerno. The Gothenburg system will be designed using its own central controller and UTOPIA SPOT units.

Theoretical and Practical Soundness

Central to the philosophy of the UTOPIA system is the provision of priority to selected public transport vehicles at signalized intersections and improvements in mobility for private vehicles. The system aims to minimize delays to private vehicles, subject to any delays necessary to accommodate priority vehicles.

The aims of the system are as follows:

(i) to provide good traffic signal coordination

(ii) to improve the flow of private traffic

(iii) to give priority to selected public transport routes by forecasting the progress of individual vehicles

(iv) to permit high flexibility in signal settings from a base reference plan

(v) to provide information (through suitable displays) to users waiting at stations about the arrival of the next coming public transport vehicle

(vi) to achieve high reliability of the whole system, including self-diagnosis, and automatically inform the central control room about failures at any point of the traffic signal network under control

Although UTOPIA is based on sound theoretical principles it has yet to be applied in a sufficient number of areas to prove its applicability in all situations.

Graphics Capabilities

Graphics facilities are only available for displaying the operation of the local controllers. These displays show the operation of the UTOPIA model and the variation in various system parameters.

System Evaluation

It is believed that research to asses the benefits of UTOPIA have not been carried out against fixed time control. The improvements attributed to UTOPIA in Turin are believed to have been calculated against some other control strategy previously installed there. Trials on the Turin network were carried out over many months. After implementing UTOPIA, private traffic speeds were found to increase 9.5 percent in 1985 and 15.9 percent in 1986, following system tuning. In peak times the speed increases were 35 percent. Public transport vehicles, which were given absolute priority, showed a speed increase of 19.9 percent in 1985.

System Setup - Parameter Selection and Validation

In addition to the specification of basic data UTOPIA requires some specification of main routes.

Ease of Access to Setup Experience

The only present application of UTOPIA is in Turin and so experience of system setup is presently limited to MIZAR and the Turin City authority. With the installation of additional systems, more information about the nature of the setup task may become apparent. Assistance and advice from MIZAR is likely to be available, but if this were required on a large scale this would probably need to be on a consultancy basis.

Advanced Traffic Management System Features

Public transport priority was important in UTOPIA from the initial design stage. It forms an integral part of the UTOPIA system. In Turing, public transport vehicles are equipped with a vehicle location system, which advises the UTOPIA system- of the location of all buses and trams in the network. Arrival time at the signals is predicted using sophisticated travel time prediction methods. Priority to buses and trams is given based on the current level of saturation, type of vehicle, and can also take into account the number of passengers on the vehicle.

Proprietary Rights

Proprietary rights are held by MIZAR and CSST for the central control software and by MIZAR for the local controller software.

Source Code Language

The central control software is written in FORTRAN while software for the local controllers is written in C.

Software Portability

The central control software can run on a multi-tasking PC using the Hewlett Packard RTE/A operating system. The local controllers are based on an industrial version of a PC-compatible board.

E. Miscellaneous

1. Weigh in Motion Systems

Many weigh-in-motion (WIM) scale systems utilize both magnetic loops and axle sensors connected to a microprocessor located in a roadside cabinet. Classification parameters such as length, axle spacing and dynamic axle weights as well as statistical parameters such as vehicle speeds and counts are collected and processed by the equipment at the WIM station. Of course, the various manufacturers of WIM systems have differing capabilities which should be evaluated before a system is considered for implementation.

Data applications

The data automatically collected by WIM has many applications. WIM data can be applied to:

- Facility planning and programming, collecting more than vehicle counts, WIM can be used to provide highway data on individual vehicles.
- Pavement design and rehabilitation, pavement structural damage, associated with load equivalency factors (LEFs) and related to the Equivalent Single Axle Load (ESAL) which corresponds to 18,000 lbs. on a single axle with dual tires.
- Weight regulation compliance. WIM scales can be used to identify trends and major deviations.
- Truck dimension compliance and regulatory policy development.

- Safety Analysis. Truck exposure and vehicle behavior can be analyzed from WIM data.
- Vehicle speed distribution for cars and trucks. Headway distributions.
- Traffic operation and control.
- Bridge load level analysis,

Accuracy of systems

WIM systems are often difficult to calibrate as one station may register a two axle truck several percent high while a three axle truck at the same station may register as several percent low. It appears that both roadway dynamics and calibration have an effect on the measurements. A methodology is suggested in the literature for calibrating weigh-in-motion systems and for monitoring that calibration over time. The proposed methodology can be used to determine when the calibration of a WIM system has changed by a significant amount (4% is proposed as significant) .

Demonstration project

Weigh-in-Motion systems have been developed and refined over the past forty years. The systems are still rapidly changing with the technology. The Crescent Demonstration Project, part of the Heavy Vehicle Electronic License Plate (HELP) program is being conducted along parts of Interstate 5 and Interstate 10 from British Columbia to Texas. The project includes an evaluation of WIM and other technologies used in automatic toll collection. The system which is being implemented will allow a truck to drive through the length of the project without having to stop at any weigh station or port of entry. The project should be able to provide a comparative evaluation of WIM systems when sufficient data has been collected.

Sensor technology

The literature contains an evaluation of data quality, performance, ease of use, and output format utility that was performed for systems with sensors from three technologies; capacitive weigh mat, bridge weighing, and piezoelectric cable. The evaluation is largely subjective. Findings show data quality roughly equivalent, with less favor of the quality from the piezoelectric system. The capacitive weigh mat and bridge systems were considered more portable than the permanently installed piezoelectric cable. Software considerations were found to provide the major differences between systems.

A capacitive weighmat is generally used in a transportable system. The mat is often used to screen overweight vehicles in a slow moving truck stream before a static weight check.

Bending plate technology provides a roadway sensor that is roughly six inches wide. A strain gauge is attached to the bending plate to measure deflection of the plate that is proportional to the vehicle weight.

A hydraulic load cell usually utilizes a heavy steel platform where the vehicle load is transferred by torsion arms to an oil filled load cell where a transducer allows the load to be measured. Both static and dynamic modes of operation can be utilized at this type of installation.

Piezoelectric sensors are usually composed of a powder filled coaxial cable that produces a voltage proportional to the applied pressure. The sensor cable is sealed inside an aluminum trough (about an inch wide) which is sealed into a roadway slot with a flexible epoxy compound.

Multisensor piezo WIM devices are suggested to more accurately estimate vehicle static weights than single sensor devices. Results show that increasing the number of sensors beyond five does not offer significant statistical improvement.

System implementation

Systems are available that obtain power from the AC line, solar cells and rechargeable batteries. Temperature compensation should be used with all sensor technologies, as the measurements from any sensor are affected. by the sensor temperature. Thermistors, located near the sensors, supply temperature correction information to the system processor to provide the automatic adjustment.

The local processor is usually located at a highway or pole mounted cabinet. A connection to the central system is generally provided by modem and telephone line. Communications is usually at 1200 to 9600 baud. Any standard means of serial data communication, including fiber optics, cellular telephone and packet radio should be capable of providing the connection. Data communication may be used to convey either continuous information updates, or blocks of data consisting of the last 24 hour period. When WIM systems are coupled with induction loops, vehicle classification can be automatically performed.

System accuracy

Dynamic pavement loading is considered in. the literature. It is suggested that various agencies; feel that WIM is not sufficiently accurate because it fails to duplicate the static weight, moreover a-device rarely records the same weight when tests are repeated. The Problem, is in the interpretation of dynamic WIM data. Other factors including vehicle suspension, roadway roughness tire type, and tire pressure are all factors influencing the dynamic forces between vehicle and roadway.

Sensor Type	Technology Description (Typical Temperature Range)
Bending Plate	Strain gauge bonded to steel plate. Strain is measured from applied toad. (-40 to 80°C)
Hydraulic Load Cell	Load from heavy. steel platform is transferred by torsion arms to oil filled load ceil. (-45 to +6O°C)
Capacitive Weight Mat	Sensor consists of an approximately 80 lb. rubber mat with steel, sheets attached as part of a tuned circuit in an oscillator. The frequency change of the oscillator is measured to calculate dynamic weight. (0 to 80°C)
Piezoelectric	A sensor consisting of a. section of coaxial cable filled with a powder, that produces a voltage when under pressure, is installed in a specially constructed slot in the roadway. Thin film piezo devices are also now available. (-40 to -60°C)